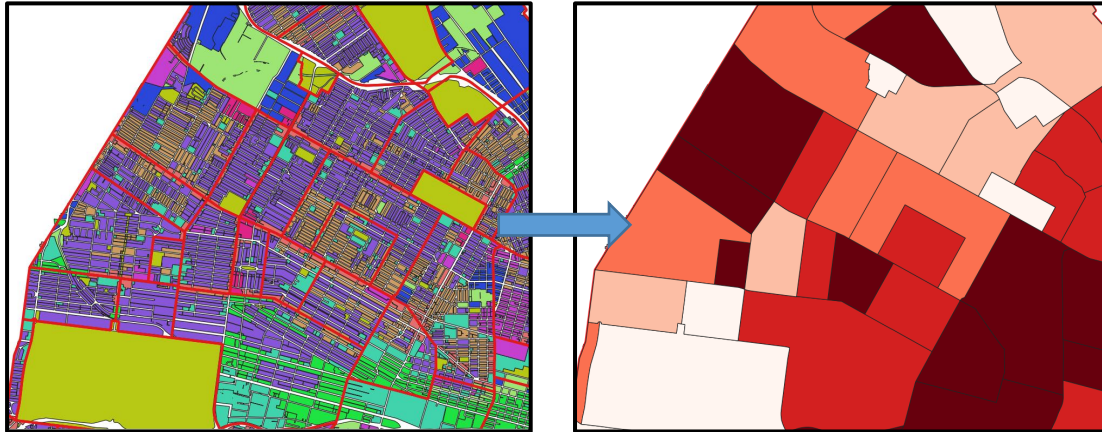# Land Use Analyzer: A Python application for Land Use/Land Cover spatial analysis

Alexandros Voukenas, GIS & Remote Sensing Engineer,
avoukenas@gmail.com

Contents

Introduction

Welcome to Land Use Analyzer, a Python application for spatial analysis of Land Use and Land Cover data! This project aims to fill in the gap on automatic implementations of some important LULC operations that are not found on most popular GIS software. Specifically, the core part of the application aims to implement indices of Land Use Mix with commonly found literature methods (mainly Entropy Index and Herfindahl–Hirschman Index). This is supplemented by other useful tools, such as extracting statistics and reports about the LULC landscape of an area.

Land Use mix is an important aspect of Urban Planning and Spatial Planning. For example, policy of mixed land use is considered an important component for promoting walkability in an urban area (Mavoa et. al., 2018) and it is considered to be energy-eficient (Zhang & Zhao, 2017). However, it was observed that no tool has been developed so far on the field of land use mix calculations, neither as standalone, nor as part of existing GIS software. Land Use Analyzer aims to fill that gap as a simple python script.

The project has a **dependency on geopandas library**. You can check here for installation options. The application is given in .ipynb format, so it should be easy to run with conda distribution and jupyter notebook.

Project structure

The following folder tree summarizes the structure of the project. It mainly consists of the land_use_mix folder, where the python files are. A folder with dummy data is also given; you can test the application with those sets of data, or you can try with your own. A temp folder is also there to store temporary data that are generated while running the application.

```
Land Use Analyzer Project                 #The project folder

|   Documentation.pdf
|   README.md
|
+---dummy_data                            #A folder with two sets of dummy data

|   +---Coastal Zones LULC                #A dataset of Coastal Zones LULC

|   |       Coastal_Zones_LC_Vector.shp
|   |       Grid.shp
|   |
|   \---SLUPShapefiles2019                #A LU dataset of San Francisco Center

|           Neighborhood_Boundaries.shp
|           SLUP2019.shp
+---land_use_mix                          #The main application folder

|   |   Land use analyzer.ipynb
|   |
\---temp                                  #A folder to store temporary data
```

Background

The application is mainly designed to calculate Land Use Mix indices. Land Use Mix indicates the level of diversity of land use types within an area. An example of different variety levels per area is given in the following figure.
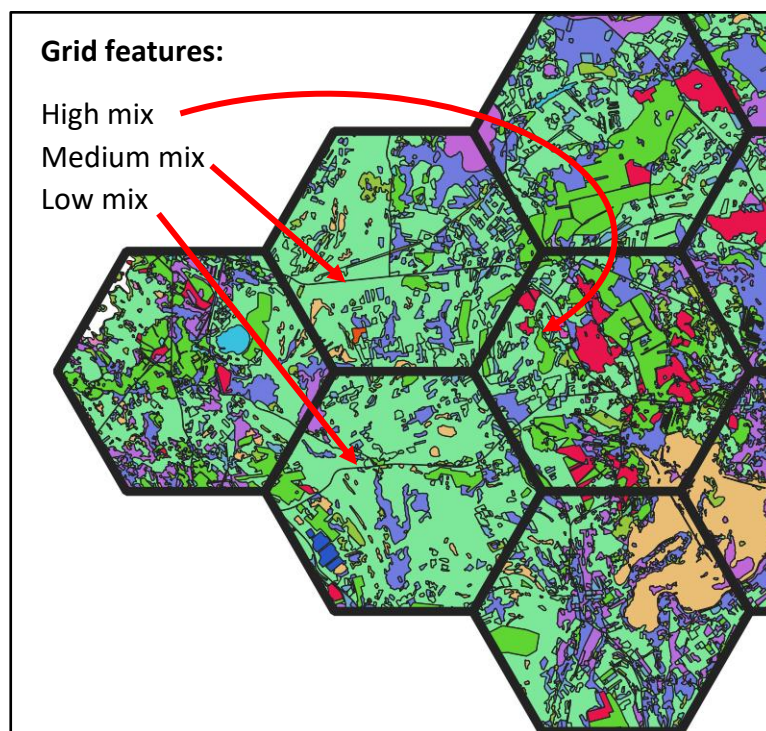
Two indices can be calculated by Land Use Analyzer, the Entropy Index and the Herfindahl-Hirschman Index (HHI). The Entropy Index is an adaptation of Shanon's Entropy:

$$Entropy = -\frac{\left(\left[\sum_{j=1}^{k} P^j \ln P^j\right]\right)}{\ln k}$$

where P^j, is the percentage of each land use type j in the area and k is the total number of land use types.

The percentage P^j is calculated in terms of area. As an example, consider an area with a total of 3 land use types:

- Residential, covering 50% of the area
- Industrial, covering 20% of the area
- Commercial, covering 30% of the area, then:

$k=3$ and $P^1 = 0.5$, $P^2 = 0.2$ and $P^3 = 0.3$

The Herfindahl-Hirschman Index is more simple, and is calculated as:

$$HHI = \sum_{j=1}^{k}(100 * P^j)^2,$$

where again P^j is the percentage of each land use type j in the area.

For further reading on Land Use Mix indices, you can consult (Song et. al., 2013)

Application analysis

To calculate the these two indices, the most important part is to calculate the percentages of each land use type. Suppose we have the same example as the screenshot, but in more detail:
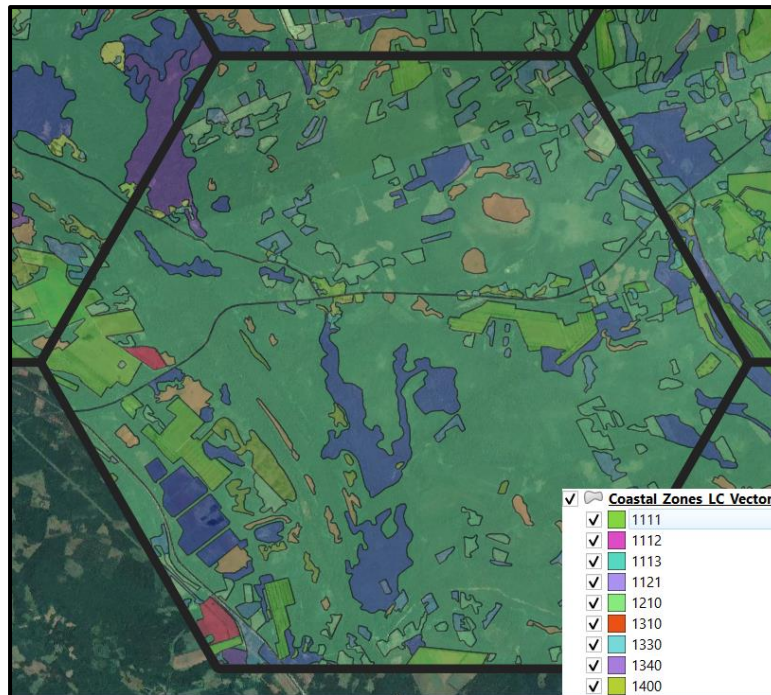
We want to calculate the Entropy Index for this feature of the hexagonal grid. The polygons inside indicate the different land use types (more correctly here, land cover types). I've also provided a legend to make more understandable how a different color corresponds to a different land use type. Each of the codes in the legend translates to a specific use (e.g., 1111 means continuous urban fabric), but it's not important to know what each one translates to.

For now, all we need to know is that each different code (1111, 1112 etc.) correspond to a different land use type. So, these codes are esentially our *j* from the forumals above. Our algorithm must follow an approach as follows:

FOR EVERY GRID FEATURE :
 *FOR EVERY CODE OF THE CONTAINED LAND USE POLYGONS:*
    CALCULATE THE SUM OF AREA OF THAT CODE
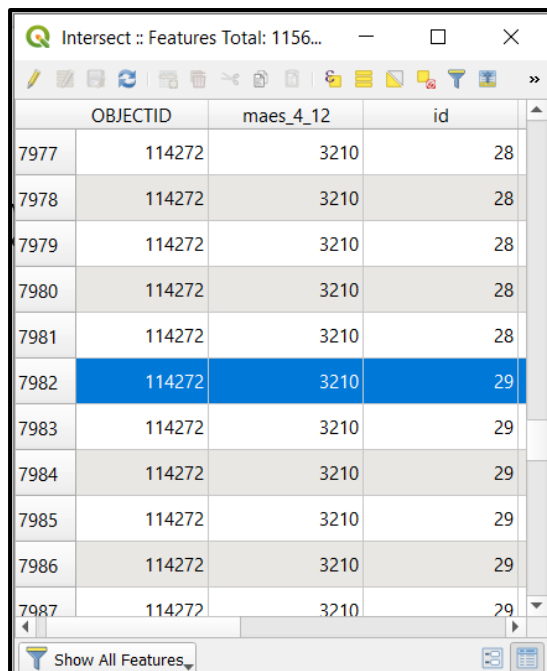   *DIVIDE IT BY THE TOTAL AREA OF THE GREAD FEATURE AND ASSIGN IT TO P^j*

After calculating the P^js it is easy to calculate the two indices. The only difficulty is to determine the land use polygons that are *contained* by each grid cell and to sum the areas of each different type. This will require two geoprocessing operations: Intersect and Dissolve. If you are familiar with GIS operations then you already understand why those are needed, but let me make it more clear with anogther figure.
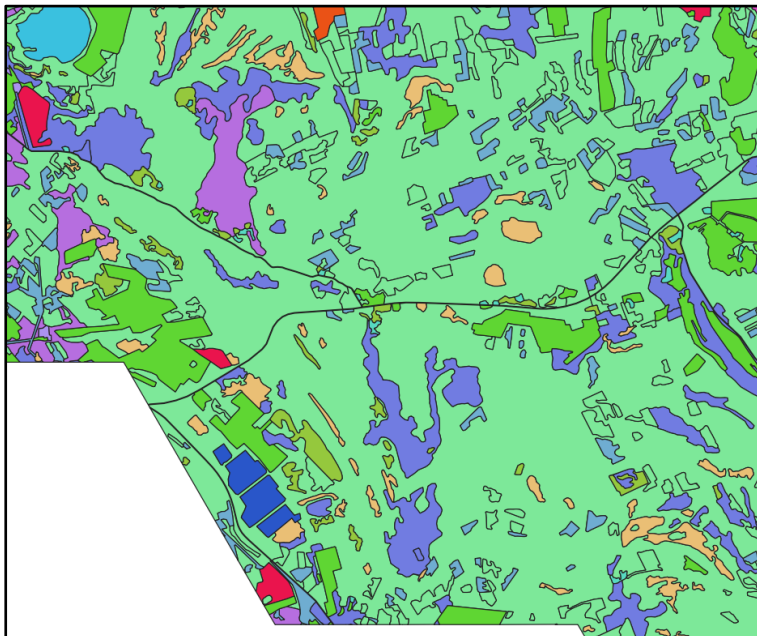
The land use dataset is continuous throughout the entire Area of Interest (throughout all the grid features). The Intersection will perform the following magic: It clips the land use polygons at the borders of each grid feature. Moreover, to each new polygon, it assigns as new attribute a unique identifier number corresponding to the grid feature it "belongs" too.
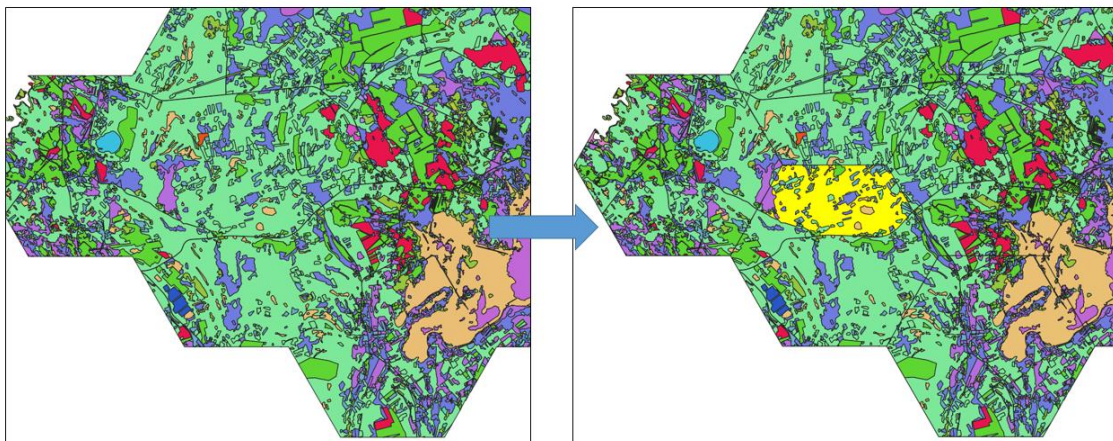
If you focus on the selected polygon (the yellow), you will see that it has been clipped to the border of the overlaying grid feature. In addition, there is a new column to the land layer which indicates the grid feature that contains the polygon. So, in our example, the selected polygon has a land use code of 3210 and it is contained by grid feature with id=29.


The Dissolve will then merge together all the polygons that have the same land use code **and** the same containing feature code. So, if we have 3 polygons with land use code 3210 and all contained under feature 29, they will become one multipolygon. And if we then calculate the area of that multipolygon, it will be the sum of the area of the code 3210 within feature 29. We divide that number by the area of the grid feature and that gives P^{3210}. For more details on how the GIS operations work, you can visit here.

In the script, this is done by function *intermediate* within the class *lum*. In the same class, the two functions to calculate the two indices (with a few more operations) are defined. The other class of the script, named *stats* contains some functions for some basic statistics. The following Table summarizes all the available functions.

Class	Function	Description
lum	*intermediate*	Performs the geoprocessing up until the calculation of P^j
	entropy	Calculates the entropy index
	hhi	Calculates the hhi
stats	*count*	For each grid feature, calculates the number of contained land use types
	mode_count	For each grid feature, identifies the contained land use type that appears most frequently
	mode_area	For each grid feature, identifies the land use type with the highest share of area
	report	Yields an overview report of the statistics of each land use type
	ratios	For each grid feature, identifies the % of cover of each contained land use type (e.g., 20% commercial, 80% residential)

For both classes, the following input must be specified:

- A feature dataset for the land use (*SLUP2019* in the following Figure)
- The name of the attribute that defines the code of each land use type (*SLUP_LATES*)
- A feature dataset that partitions the area into smaller divisions, e.g. administrative or municipal boundaries (*Neighborhood Boundaries*)
- The name of the attribute that defines a unique identifier for each division (*NHD_NUM_ST*)
- A folder directory for some temporary data to be stored
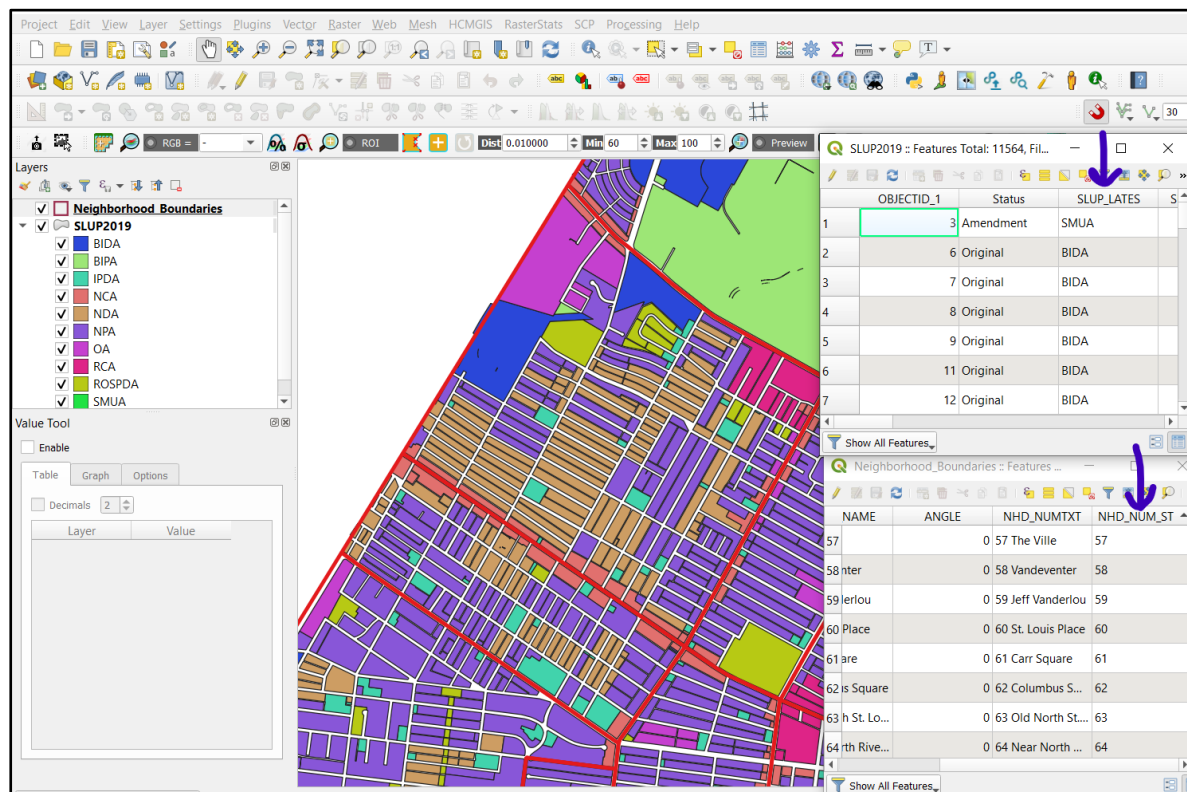
Example result

In this example, a land use dataset from the center of Saint Louis is used, along with a dataset of Saint Louis neighborhoods. The entropy index has been calculated for each neihborhood polygon and the more red the color, the higher the entropy.
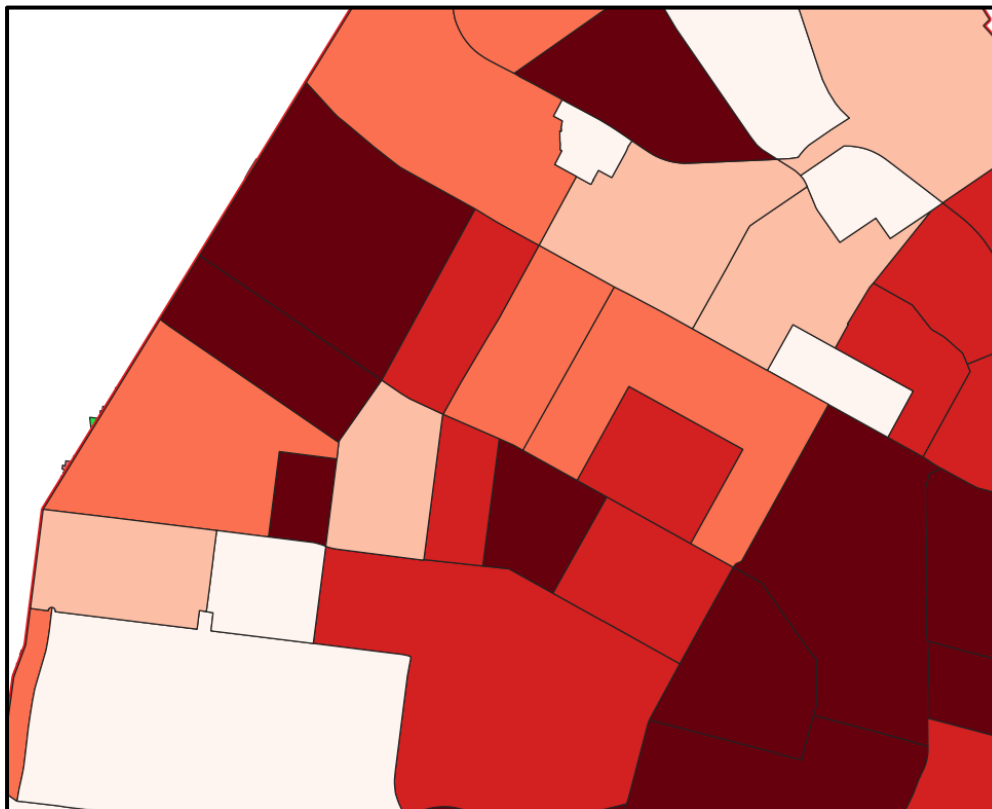
QGIS Plugin

All of those features have been incorporated into a plugin for QGIS, named Land Use Analyzer. It can be downloaded either from the plugin repository, or (the latest version), from the current repository, under folder "QGIS plugin". Download the contents of the folder as a .zip file and use the option Install from ZIP from Plugin manager in QGIS.

Once the plugin has been installed, you will see it under the Processing toolbox, as well as the Plugin Menu.
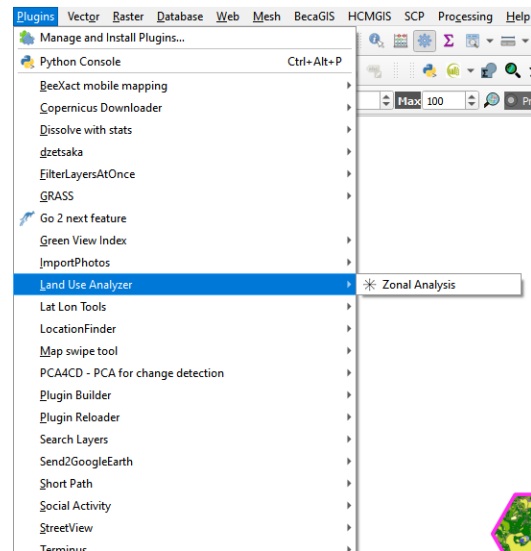
The tool "Zonal Analysis" is currently the only tool, but it includes all the analysis methods of the main project, except for the overview report.
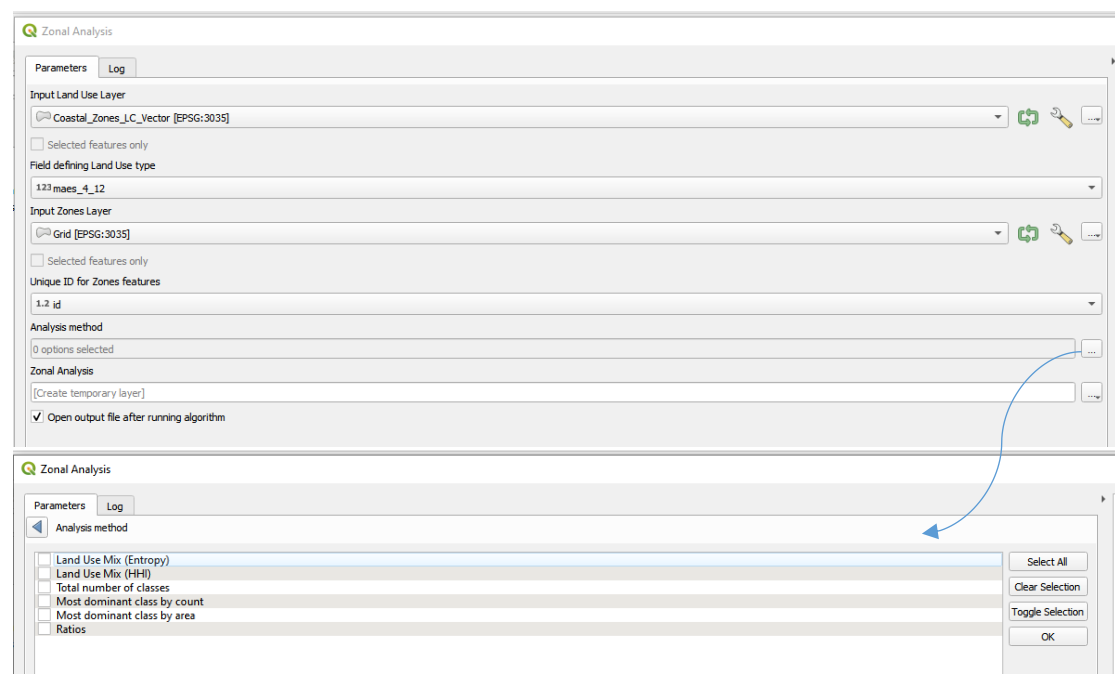
Literature

Mavoa, S. et al., 2018. Identifying appropriate land-use mix measures for use in a national walkability index. *The Journal of Transport and Land Use,* 11(1), pp. 681-700.

Song, Y., Martin, L. & Rodriguez, D., 2013. Comparing measures of urban land use mix. *Computers, Environment and Urban Systems,* Volume 42, pp. 1-13.

Zhang, M. & Zhao, P., 2017. The impact of land-use mix on residents' travel energy consumption: New evidence from Beijing. *Transportation Research Part D,* Volume 57, pp. 224-236.

Contact

Alexandros Voukenas, avoukenas@gmail.com