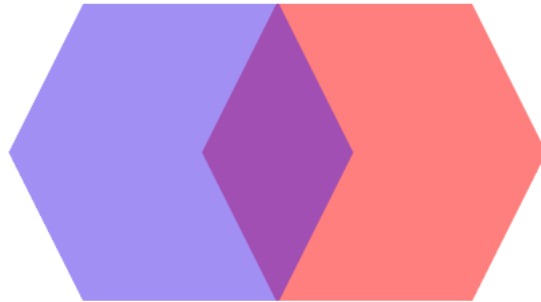


GeoCluster Plugin

User's Manual



Contents

1.	Introduction	3
2.	Procedural Steps	3
3.	Steps description.....	4
3.1	Grid Creation and layers intersection.....	4
3.2	Group Stats plugin Elaborations	5
3.3	Join Step.....	6
3.4	GeoDa clustering Elaborations	7
3.5	Clusters extraction.....	9
4.	References	10

1. Introduction

The *GeoCluster* plugin allows users to cluster different territorial scopes (N° of clusters = 2). To correctly use *GeoCluster* you need to know that:

- i. All the analyses are intended for vector layers only. To ensure proper plugin functionality, the selected shapefile must contain a field named "code" (text format). The latter should contain the information used for clustering purposes. In the next steps, clustering operations will be performed based on specific land use land cover (LULC). Here, the field "code" contains the type of LULC associated with each polygon (retrieved from the Urban Atlas dataset). In this case, values ranged from 1 to 5.
- ii. The plugin is based on the probability space represented by a hexagonal mesh grid, shifted in its main directions (vertical, horizontal, and diagonal). By considering the individual shifts, the clustering results will be merged, obtaining the final cluster groups. For further information see [Fiorini et al., 2022](#).
- iii. The coordinate reference system must be in metric units (EPSG 32633 or EPSG 32632).
- iv. This plugin is intended to be used together with the GeoDa Software (or the "Attribute-based clustering" plugin) and the GroupStat plugin. Specifically, between step 1 and step 2 clustering operations will be required. The use of GeoDa is not mandatory. However, field header must be respected as described in this guide.
- v. Discretization results should always be checked before proceeding to further analyses.

2. Procedural Steps

The GeoCluster plugin articulates into the subsequent analytical steps:

1. Grid creation - The plugin extracts a hexagonal mesh grid (mesh area of 1 hectare), considering the extension of the provided layer. The grid is then shifted horizontally, vertically, and diagonally. Finally, to each polygon the mesh ID within which it falls is associated.

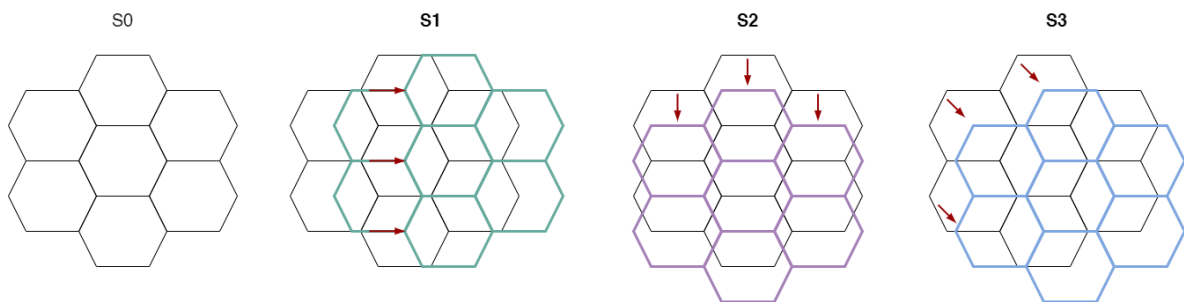


Figure 1. Grid shifts

2. GroupStat elaborations: This plugin is used to calculate the areas of each polygon, based on the "code" field. The results are then saved in CSV format.
3. Join: Information about the areas calculated through the GroupStat plugin are associated to the grid meshes, through their ID.

4. Clustering elaborations: A Cluster group is associated to each mesh. For this step GeoDa software can be used.
5. Cluster extraction: Clusters associated to each grid mesh are extracted and merged, based on the group they belong. The algorithm passes from 4 layers (one for each shift) to 2 layers (one for each cluster group). Finally, the overlapping portions of the final cluster groups are extracted.

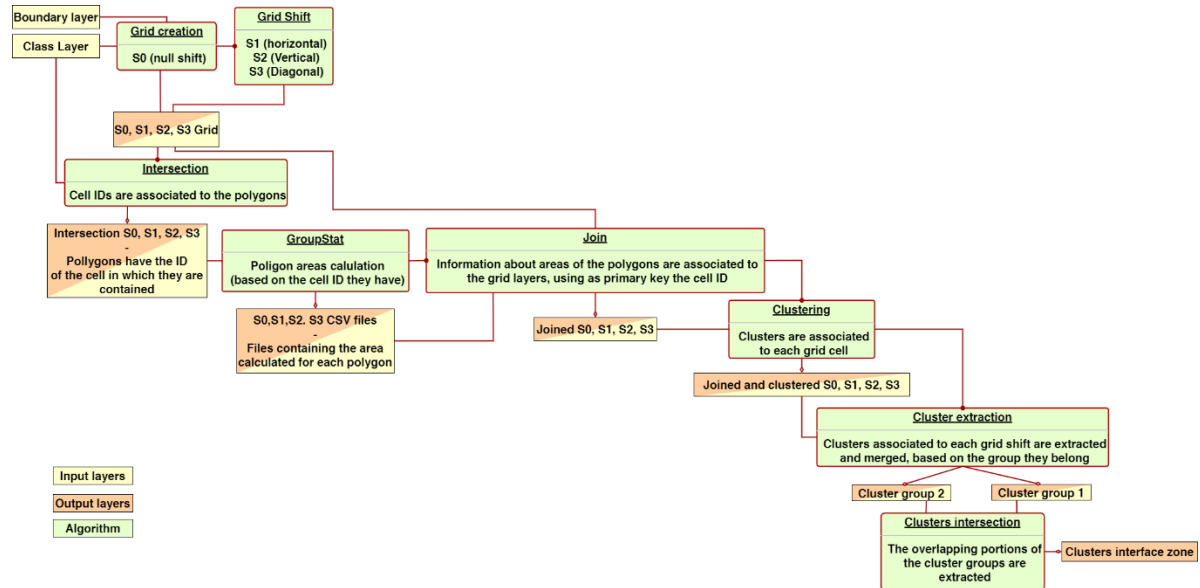


Figure 2. Workflow structure

3. Steps description

3.1 Grid Creation and layers intersection

INPUT

1. Class layer - Layer containing the desired categories, reported into the “code” field.
2. Class layer boundary – Layer containing only the boundaries of the study area.

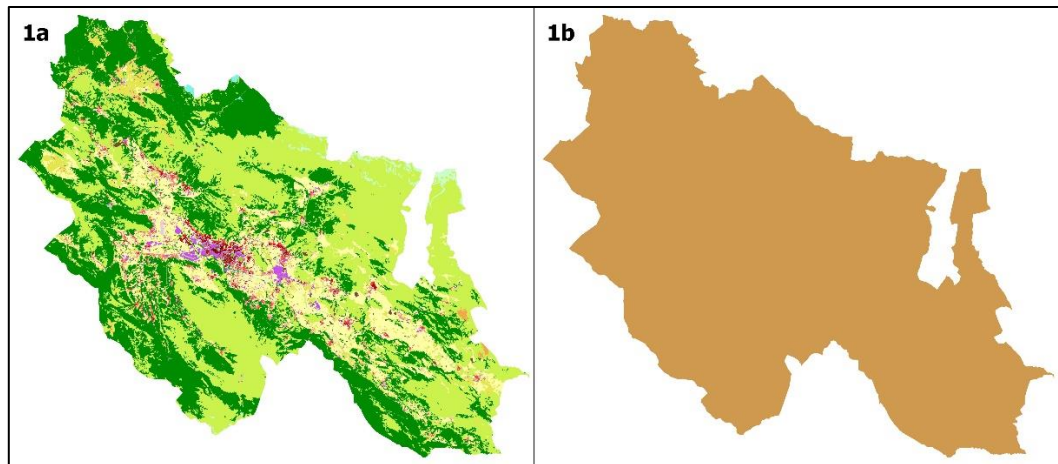


Figure 3. Class layer (1a) and Class layer boundary (1b).

OUTPUT

1. **Intersection (Can be S_0, S_1, S_2 or S_3 indicating the single grid shift)** – Layers obtained from the intersection between the Grid shifts and the Class layer. Intersection S_0, S_1, S_2 or S_3 layers will contain the information retrieved from the grid layer (cell id) and from the Class Layer.
2. **(S_0, S_1, S_2, S_3) Grid** – The grid layers that will be used in the second step (Join step)

3.2 Group Stats plugin Elaborations

Starting from the S_0, S_1, S_2 or S_3 intersections, the Group Stats plugin is used to calculate the areas inside each tile, based on the “code” field and the tile’s ID. Set up the values as in Table 1 and Figure 4, then make sure to save the relative CSV (Data > Save all to CSV file).

LAYER	COLUMNS	ROWS	VALUE
Intersection (S_0, S_1, S_2, S_3)	code	ID_	Sum & Area

Table 1. Group Stats boxes and values to be used.

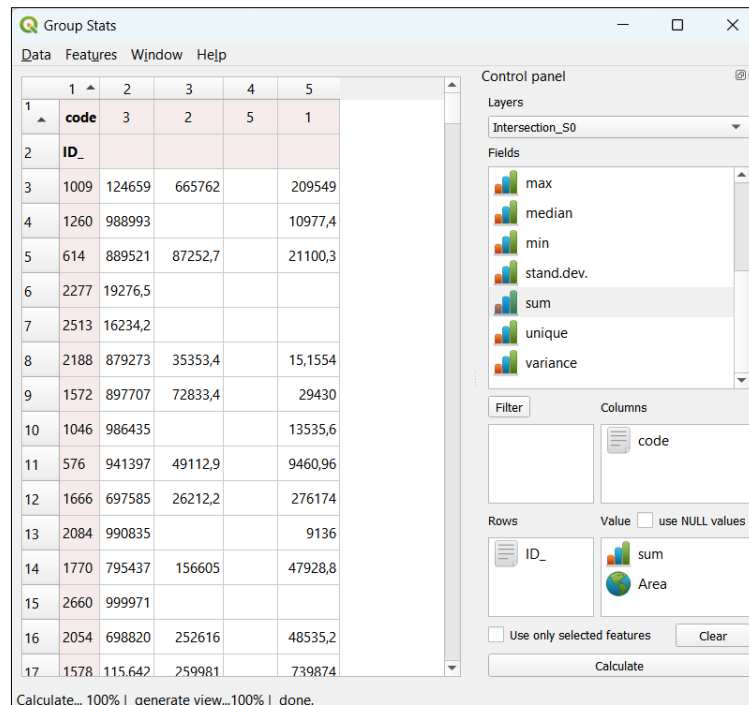


Figure 4. Group Stats user's interface. *N.B.* When selecting values in the relative box, make sure to select the correct ones (the "world" icon, not eventual "pie chart" icon values, and the "fact sheet" icons).

3.3 Join Step

In the Join step the newly created CSV columns will be added to the Grid files through the join process. Tiles information will be joined through their ID.

INPUT

1. S_0, S_1, S_2, S_3 CSV: CSV files previously saved from the Group Stat plugin.
2. S_0, S_1, S_2, S_3 Grid: Files obtained from the "Grid Creation and layers intersection" step.

OUTPUT

1. Joined S_0, S_1, S_2, S_3

Import all the four CSV files as shown in Figure 5. Then, fill each field with the corresponding file in the plugin step window.

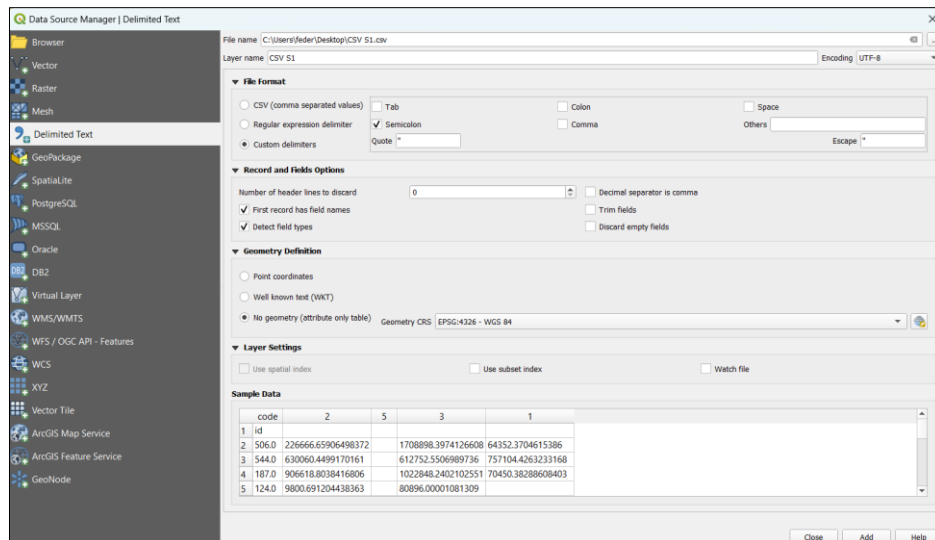


Figure 5. CSV import procedure (to be repeated for all the four CSVs).

Right click on the “S_n Grid” layer, then select the “Join” procedure. Click on the “+” button and set all the variables as shown in Figure 6.

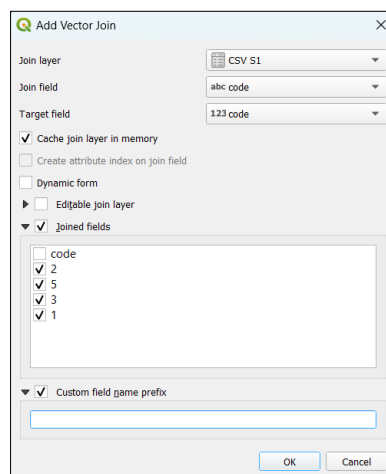


Figure 6. Join procedure (to be repeated for all the four CSVs).

3.4 GeoDa clustering Elaborations

Although being realized through GeoDa, the clustering phase can be realized with any software or plugin. Regardless the mode of clustering, the column in the attribute table corresponding to the clusters must be named as “kmeans”. Otherwise, next steps won’t work.

INPUT

Joined S₀, S₁, S₂, S₃: Layer files with all the fields from CSV files joined, for each grid cell.

OUTPUT

Joined and clustered S₀, S₁, S₂, S₃: Layer file with all the fields from CSV files joined, for each grid cell. These new files will include the “kmeans” field, reporting the cluster associated to every grid cell.

Import the “joined S_n ” shapefile (output from the join step, i.e. joined S_0 , joined S_1 , joined S_2 , joined S_3) inside the GeoDa software. Then, by selecting the clusters voice in the tools bar, choose the Kmeans method (Figure 7) and fill in the fields as shown in Figure 8. Run the algorithm and then save the shapefile.

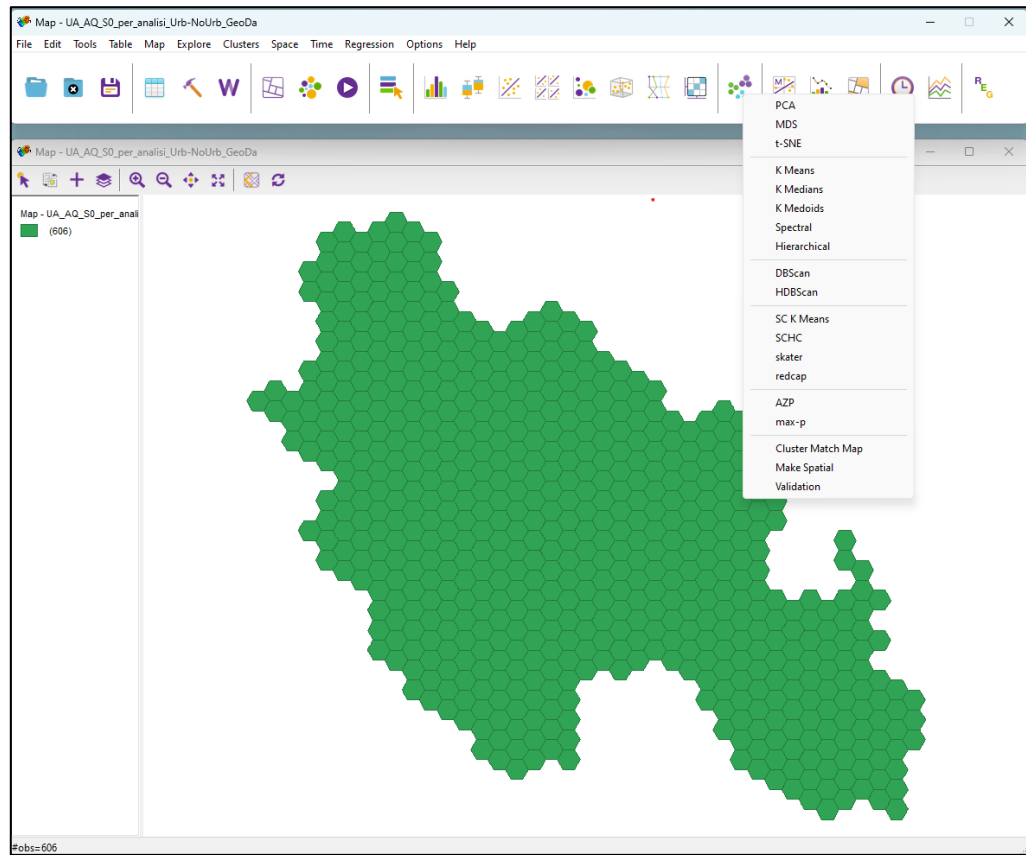


Figure 7. Clustering window selection (GeoDa software)

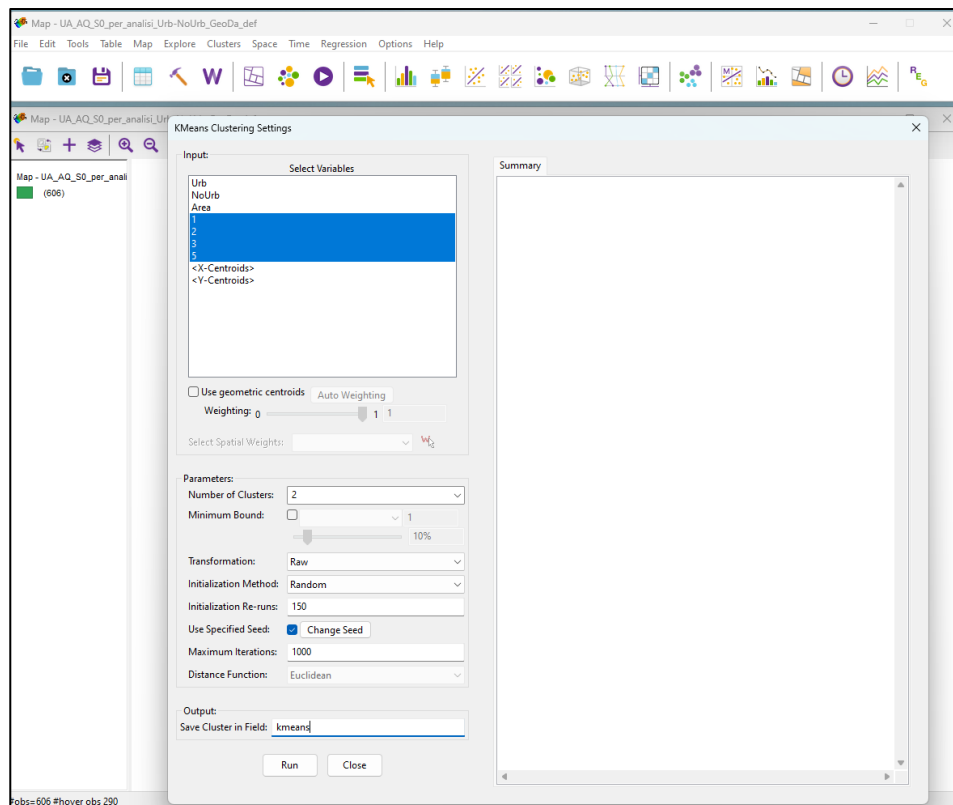


Figure 8. “Kmeans” clustering method field filling. N.B. In the “Selected variables” box only 4 variables have been selected. This is a specific situation in which only four types of land cover are present. By following all the procedures, the maximum number of variables is five, corresponding to the numbers 1,2,3,4 and 5). For more information, please consult Fiorini, L., Falasca, F., Marucci, A., & Saganeiti, L. (2022). Discretization of the Urban and Non-Urban Shape: Unsupervised Machine Learning Techniques for Territorial Planning. *Applied Sciences*, 12(20), 10439.

3.5 Clusters extraction

Once all the clustering operations have been realized, the last feature extraction step can be run. Here clusters will be extracted, highlighting their overlapping portions.

INPUT

Joined 0, 1, 2, 3: Joined and clustered shapefiles, resulting from the GeoDa software.

OUTPUT

Cluster group extraction 1/2: Clustering group 1 extraction

Cluster group extraction 2/2: Clustering group 2 extraction

Clusters interface zone: Overlapping portions derived by Group 1 and Group 2 Intersection

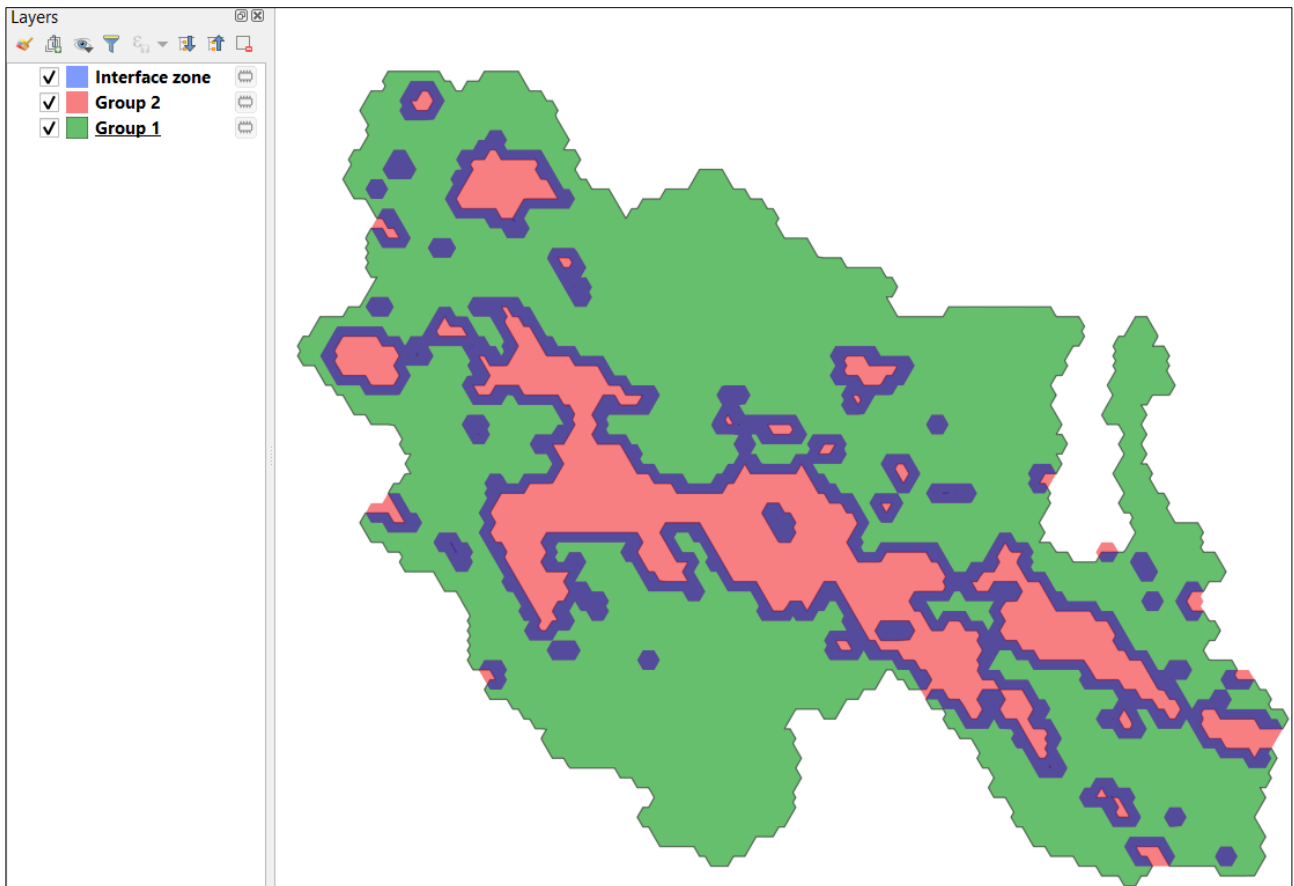


Figure 9. Clusters extraction, final results.

4. References

This plugin is one of the outcomes of the paper entitled “Discretization of the Urban and Non-Urban Shape: Unsupervised Machine Learning Techniques for Territorial Planning” (<https://www.mdpi.com/2076-3417/12/20/10439>). If you use this plugin for your elaborations, please consider citing us:

Citation: Fiorini, L., Falasca, F., Marucci, A., & Saganeiti, L. (2022). Discretization of the Urban and Non-Urban Shape: Unsupervised Machine Learning Techniques for Territorial Planning. *Applied Sciences*, 12(20), 10439.

FOR ANY TROUBLESHOOTING DO NOT HESITATE TO CONTACT US

MAIL: FEDERICO.FALASCA@GRADUATE.UNIVAQ.IT