SIE-167 - Introduction to Environmental Engineering - Exercise
Stéphane Joost, Noé Fellay, Gabriel Kathari

Environmental characterization of inhabited hectares in the canton of Geneva

Objective

The goal of this exercise is to create a dataset containing environmental descriptors at a hectometric grid level. You will then reproduce the same approach using the GIREC statistical subdivisions of Geneva (découpage en sous-secteurs statistiques à Genève) and build thematic maps in QGIS. Additionally, you will carry out exploratory data analysis using GeoDa software.

If you did not install QGIS and Geoda before, here are the links:

- QGIS, version 3.34 LTR → https://www.qgis.org/download/
- Geoda, version 1.22.08 → https://geodacenter.github.io/download.html

In Moodle (https://moodle.epfl.ch/course/view.php?id=501#section-6) you will be able to download a .zip file (GIS data.zip). After downloading, extract the following datasets:

1. **Land Surface Temperature** (LST) is the mean for months of June, July and August in 2009, 2010 and 2011 using Landsat 7 images. The same projection system (Swiss coordinate system LV03, EPSG 21781) is used for all the datasets.

2. **Normalized Difference Vegetation Index** (NDVI) is the mean for months of June, July and August in 2009, 2010 and 2011 using Landsat 7 images.

   These data were extracted using Google Earth Engine.

   Information on Landsat bands can be found here: https://www.usgs.gov/faqs/what-are-best-landsat-spectral-bands-use-my-research?

3. SonBase **road traffic noise** produced by the Swiss Federal Office for the Environment, characterizing noise during **day** time and **night** time. You can read the documentation hereunder to understand how this database was produced: https://www.bafu.admin.ch/bafu/en/home/topics/noise/state/gis-laermdatenbank-sonbase.html

4. Air pollution **N02, PM10** and **PM2.5** produced by the Swiss Federal Office for the Environment and the Meteotest company. See following link: Map of annual values

5. Hectometric vector grid ("hecto.grid.ge.2015") representing inhabited hectares in the Canton of Geneva in 2015. Produced by the Swiss Federal Statistical Office, GEOSTAT service, the center of competence for geoinformation and digital image processing:
   https://www.bfs.admin.ch/bfs/en/home/services/geostat.html

6. **The statistical sectors in Geneva (GIRECs)** (« GEO_GIREC ») **:**
   - See a description in the SITG (Système d'Information du Territoire Genevois) catalogue: Statistical subdivisions of Geneva.

Tasks

Part 1: Creating an environmental dataset at the hectometric grid level

Open QGIS and check the language of the application. If the menus and toolbars are not in English, you need to switch to this language in order to follow the instructions. Hereunder we refer to the English terms so that everybody understands, but the name of the menus, and menu items will be in the language of the OS.

1. In the menu "Settings" ("Préférences") select the line "Options" (idem in french);
2. In the column on the left, select the first line "General" ("Général")
3. In the right part, check the first box on the top "Override System Locale" ("Remplacer les paramètres régionaux du système")
4. On the next line "User interface translation" ("Traduction de l'interface utilisateur") select "American English" and click OK.
5. Close QGIS and open it again (restart of QGIS). Then you are ready.

1. **Download and prepare data**
   - Download the files from Moodle and save them to a local directory. Launch QGIS and create a new project.

2. **Import raster layers**
   - Import the LST, NDVI, the 2 SonBase road traffic noise (night and day), and the 3 air pollution files (NO2, PM2.5 and PM10).

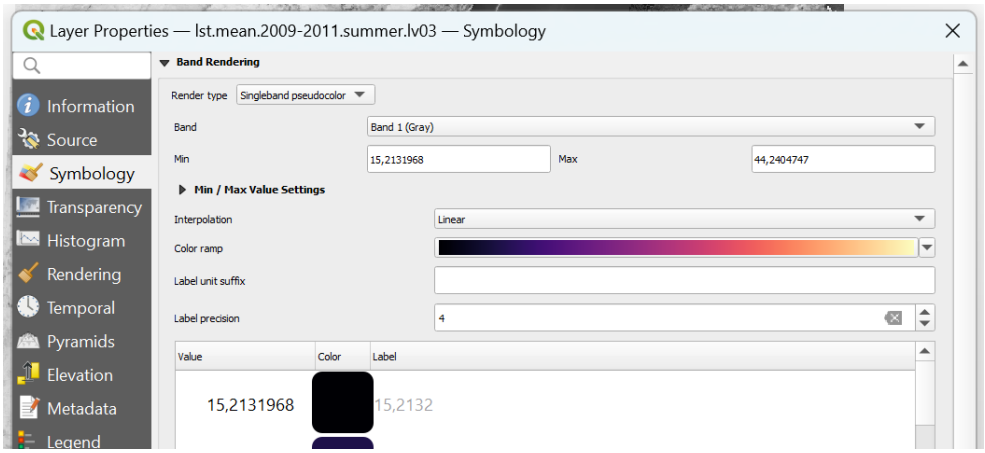
   Steps:
   - Go to **Layer** -> **Add Layer** -> **Add Raster Layer** -> 
   - Select all raster files and click **Add**.
   - Click **Add**.

3. **Import vector layer**
    - Import the hectometric grid "**hecto.grid.ge.2015**.

      Steps:
    - Go to **Layer -> Add Layer -> Add Vector Layer ->** 
    - Select hecto.grid.ge.2015.shp and click **Add**.

4. **Explore the LST raster**

   Visualization:



    - Right-click on lst.mean.2009-2011.summer.lv03 → **Properties** → **Symbology**.
    - Set **Render type** to **Singleband pseudocolor**.Under **Color ramp,** select **Magma**. Click **Apply** and **OK**

   Question 1: Do you notice any spatial pattern? Where do the highest and lowest surface temperatures appear?

5. **Analyze the LST histogram**
    - Right click on **lst.mean.2009-2011.summer.lv03** -> **Properties** -> **Histogram** Click **Compute Histogram**.

      Question 2: In what range values are the most frequent land surface temperature values?

6. **Noise and pollution**
    - Observe the noise rasters and pollution rasters on QGIS.

      Question 3: Do you notice any similarities between the sonbase.gva.day.2015 and pm2.5.mean.2015.lv03 rasters? Why might these rasters be similar ?

7. **Perform zonal statistics to transfer raster data to the hectometric grid**

   Purpose: Calculate statistics in the hectometric cells (e.g. mean LST, mean NO2) using the information stored in the raster layers.

   Steps:

- Open the **Processing Toolbox**: **Processing** → **Toolbox**.
- Search for **"Zonal statistics"** and double-click on it.
- **Input layer:** Select **hecto.grid.ge.2015**.
- **Raster layer:** Select one of the raster files (start with the NDVI raster).
- Click on the three dots next to **"Statistics to calculate"** and select **Mean**, **Median**, and **Standard deviation**.
- **Output column prefix:** Use a relevant prefix (e.g., for NDVI, use **“ndvi_”**).

  [Ensure the prefix plus statistic name does not exceed 10 characters]

- **Run** the tool.

After the computation of the first zonal statistics, a layer named **“Zonal Statistics”** will appear in the Layers panel on the left, with 3 new added columns (you can check with the menu Layer > Open Attributes Table, and look at the last columns on the right). Now, run the zonal statistics tool again.

- **Input layer:** Use the **updated layer** that includes the previous zonal statistics (the output from the last step).
- Repeat the process for each of the remaining raster layers (LST, SonBase day/night, NO_2, PM2.5, PM10), ensuring you always use the latest updated layer as the input (i.e. the latest version of the “Zonal Statistics” file).

[This ensures all statistics are accumulated in a single layer.]

After processing all rasters, right-click on the final zonal statistics layer and open the **Attribute Table**. Check that the 21 new columns have been recorded in the target layer.

Then save your final zonal statistics layer in your folder.

- Right-click on the layer -> **Export** -> **Save Features As** -> Click on the triple dots -> Save as **hecto.grid.ge.2015_final.shp**. Tick the box to display this version of the file in your QGIS project.

8. **Create a thematic map of NDVI mean**

   Steps:

   - Right-click on hecto.grid.ge.2015_final.shp → **Properties** → **Symbology**. **Set Render type** to **Graduated**.
   - **Value:** Select **ndvi_mean.**
   - **Classification Mode**: Choose **Natural Breaks (Jenks)**.
   - **Classes**: Set to 5.
   - **Color Ramp:** Select a gradient from light green (low NDVI) to dark green (high NDVI).High NDVI values should be represented with a **dark green** hue, and low NDVI values with a **light green** hue.
   - Click **Classify**, then **Apply,** and **OK**.

   Question 4: Which parts of Geneva have the lowest NDVI values? Now, do a similar map with the value of your choice (noise, pollution or land surface temperature). Change the classification mode and color ramp and create a new thematic map.

Part 2: Building a thematic map using GIREC shapefile

You will now build a thematic map using the GIREC shapefile and PM2.5 data. Download the **"GEO_GIREC.shp"** file from Moodle and add it to your QGIS project.

1. **Convert PM2.5 raster:**

   Since we want the PM2.5 to be expressed in µg/m³ instead of dg/m³, we will convert it using the Raster Calculator.

   - Go to **Raster** -> **Raster calculator**.
   - **Raster Calculator Expression:** "pm25.mean.2015.lv03@1" / 10 (Converts dg/m³ to µg/m³ (1 dg = 10 µg).
   - **Set the Output Layer:**
   - Click on the **…** (three dots) next to **Output layer**.
   - **Output Layer:** Save as pm25_mean_µgm3.tif.
   - Click **OK**.

2. **Perform zonal statistics on GIREC shapefile:**

   - According to previous instructions, **compute the mean $PM_{2.5}$ concentration** on the "**GEO_GIREC**" shapefile using the zonal statistics.

3. **following the instructions in Question 5**

Question 5: Create a thematic map in 5 classes using the GEO_GIREC layer according to previous instructions (Part 1, task 8). Select a graduated classification with a color ramp of your choice on the value **pm25_mean**.

.

4. **Create a thematic map within the map layout:**

    - Go to **Project** -> **New Print Layout**.
    - Name it (e.g., PM25_Map_Layout).

Add map elements:

1. **Map frame:**

- **Add Item** -> **Add Map**.
- Draw a rectangle through the page, in which the map will appear (draw through the entire page).

2. **Title:**

- **Add Item** -> **Add Label**.
- Set the title: "Mean $PM_{2.5}$ concentration by GIREC sector in Geneva" at the top of the Map.
- In the **Item Properties** panel, under **Main properties**, adjust font size and alignment. Put the title in Arial, Bold, 16 pt and center it.

3. **Legend**:

- Add Item -> Add Legend
- You can customize the legend in the Item Properties panel to display only relevant layers. For example:
    - **Legend Items**: click on “Only show items inside linked map” -> click on “Auto update”.
    - Rename Legend Entries: Unclick “Auto-update” and double-click on the layer name under “Legend Items” (e.g., “Zonal statistics”) and rename it to “**$PM_{2.5}$ Mean (µg/m³)**” for clarity.
    -

4. **Scale bar** and **North Arrow**:

- Add Item -> Add Scale Bar / Add North Arrow -> Add them in the bottom right of your map frame (you can try to customize the scale bar and north arrow if you want).

5. **Additional Labels**:

- Use the Add Label tool to insert your name, date and data sources in the bottom left of the map.

Note: for those who are motivated, you can overlay an “OpenStreetMap” background to the QGIS layers. For that you need to switch back to QGIS (QGIS and the Print Layout are in two separated windows) and then in the Browser panel

go to XYZ Tiles -> OpenStreetMap. You can drag the OSM layer to the Layer panel and place it behind the GIREC layer.

6. **Export the Map**
   - Go to Layout -> Export as Image
   - Save as PM25_GIREC_MAP.png

An example of expected output is represented in the Figure 1 below:

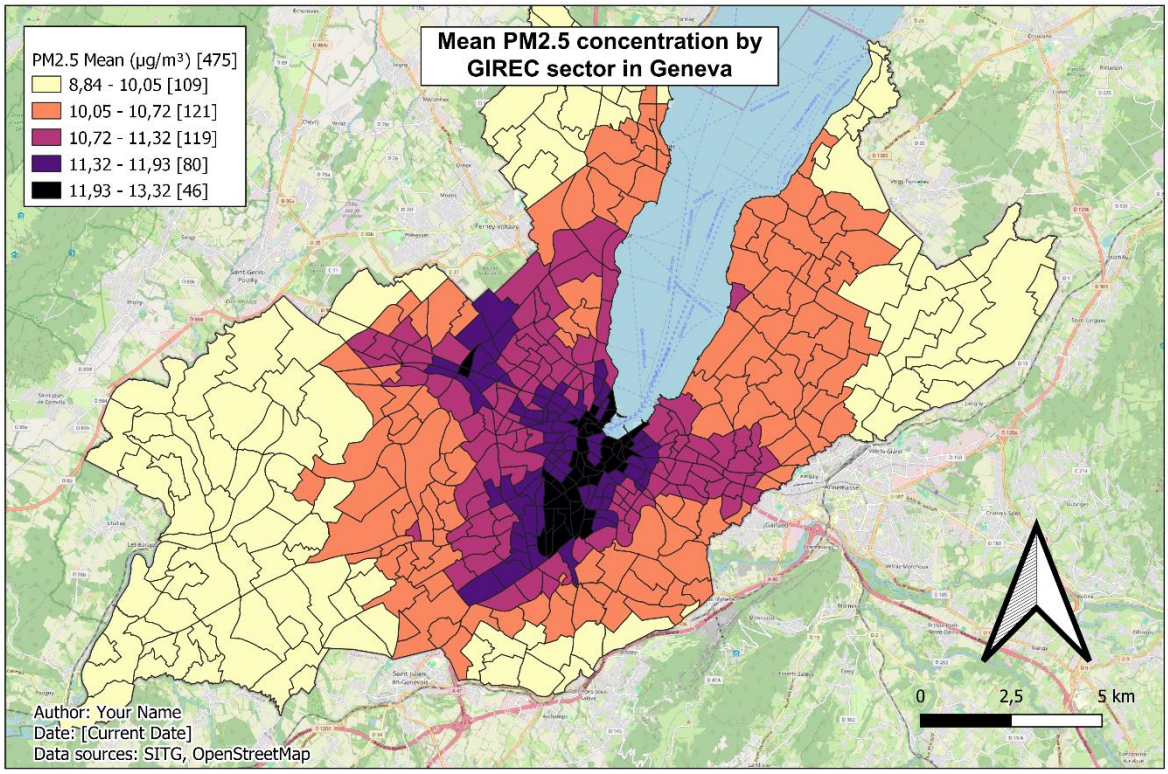


Figure 1 : Mean PM2.5 concentration by GIREC sector in Geneva.

Question 6: Which area shows the highest mean $PM_{2.5}$ levels? What could the reason for those high concentrations?

Question 7: What spatial patterns do you observe in the distribution of $PM_{2.5}$ concentrations across the GIREC sectors? Do they seem to be randomly distributed? Do you notice any similarities with the thematic map you did on NDVI in Part 1 task 8? Do you think there is a link between NDVI and $PM_{2.5}$?

Part 3: Introduction to GeoDa

1. **Launch GeoDa and load the data you created in Part 1.**
   - Drag and drop “hecto.grid.ge.2015_final.shp” into GeoDa.

2. **Exploring data distributions:**

Task 1: Histogram of nighttime noise levels

- Go to Explore -> Histogram.
- Select the variable for nighttime noise levels (e.g., “rdtn_mean”). Click Ok.
- Right-click on the histogram -> **View** -> **Display Statistics**.
- Keep the Themeless Map open (Map menu, first line). Place the map window and the histogram window one next to the other.

Question 10: What is a histogram, and what does it tell you about nighttime noise levels? Describe the distribution shape (skewed, normal). What does this tell you about the noise levels across different areas? Provide the min, max and median.

3. **Identifying outliers with boxplots:**

Task 2: Create a boxplot of NDVI

- Go to Explore -> Box Plot.
- Select ndvi_mean:
- Observe the quartiles, median, and identify any outliers. (Right click on histogram -> view -> display statistics.

Question 8: Describe the boxplot: min, Q1, median, mean, Q3, max. On the boxplot, what are the black lines, the orange line, the green dot and the red box?

Question 9: Are there any lower outliers? If so, how many, and to which part of Geneva do they correspond? (**Hint**: you can select the outliers in a similar manner like on Figure 2 and check where they fall on the map).

Figure 2 : Example on how you can select variables on GeoDa.

4. Investigating relationships between variables:

Task 3: Scatter plots

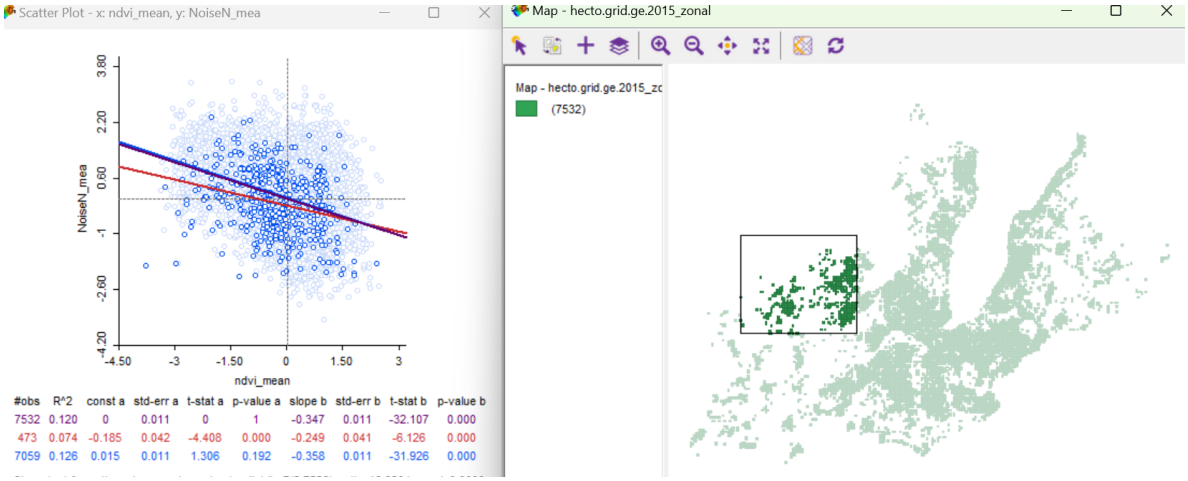
- Mean **NDVI** vs. mean **$PM_{2.5}$**

- Go to **Explore → Scatter Plot**.
- Set X Variable to ndvi_mean.
- Set Y Variable to pm25_mean.
- Click Ok.

Question 10: Is there a significant relationship between NDVI and $PM_{2.5}$? (Hint: look at the p-value). What does the slope indicate?

Question 11: Repeat with the mean noise during the night as Y variable.
Is there a significant relationship between NDVI and nighttime noise levels? Discuss possible reasons.

Using GeoDa, you can draw a rectangle on the Map as shown below. By holding your mouse on the rectangle and moving, you can see how the relationships between the X and Y variables change across the territory.



Question 12: Is the relationship between NDVI and noise uniform across the territory? How can you see that on the scatter plot?