

# Geographic Information Systems 2023-2024

---

## Assignment 01

---

Date due: 28th of May 2024, 8pm.

### Context

- This assignment will contribute 8/20 (maximum) to the final grade on the SIG/GIS course.
- The work may be undersigned by a group of two students (maximum) and will be discussed individually.
- The group members should be informed to the teacher on the class of 30th of April, or by e-mail to [ruifigueira@isa.ulisboa.pt](mailto:ruifigueira@isa.ulisboa.pt) before 30th of April 2024, 8pm. Include ISA's numbers, names and e-mails. Regarding groups with 2 students, one e-mail is enough provided it includes in cc the other student.
- Each region/polygon in Problem 02 will specify to each group (**identified in the attribute *StudentGroup* of the layer *EditingZones***), and will be randomly assigned by e-mail before class of 30th of April 2024.
- **The work should be submitted before the deadline through the form <https://forms.gle/3AuWQVhj2mtSeU5A>.**
- **The deadline to submit the full work is 28th of May 2024, 8pm.**

### Evaluation criteria

- Correctness of solutions and answers – 50%
- Methods used on problem solving – 20%
- Clarity and critic view of the report – 20%
- Coherence of the overall work – 10%

### Deliverables

- A zip file of the main project directory (Problems 01 to 04) containing
  - The *project file* containing all the output and relevant intermediary layers (legends and labels when required and/or appropriate)
  - The subdirectories *DataIn* and *DataOut*
  - All the final data sets that answer the questions (using the most appropriate file formats)
  - All the relevant initial intermediary data sets used to reach the final answers (using the most appropriate file formats)
  - Because of file and storage size, **do not include** in the project the original DEM downloaded from <http://earthexplorer.usgs.gov/> to solve problem 01.
- A zip file of the additional project directory (Problem 05) containing
  - The *project file* containing all the output and relevant intermediary layers (legends and labels when required and/or appropriate)
  - The subdirectories *DataIn* and *DataOut*

- All the final data sets that answer the questions (using the most appropriate file formats)
- All the relevant initial intermediary data sets used to reach the final answers (using the most appropriate file formats)
- A report (pdf file format) with 3 pages (maximum! using font 12 and single line spacing paragraphs) explaining the adopted solution for every question (references to diagrams are allowed and strongly recommended) and the main theoretical, conceptual and technical issues underlying the choice of that solution. Concerning the implemented processes to solve the problems, include in the report only those details that are impossible to represent in diagrams. Do not include in the report details related to map legends and/or labels.
- pdf files (as many as needed) containing the diagrams that represent the implemented solutions when required (use in diagrams the names of the layers included in the zip file).

## Input data

- The zip file [Assignment2024.zip](#) available at [FENIX](#) assignment page containing the following files:
  - [Territory.gpkg](#) (geoPackage format) containing the following spatial data sets:
    - River
    - GwPollVulnRegion
    - SoilType
  - [Parcels2024.gpkg](#) (geoPackage format) containing the following spatial data sets:
    - Cadastre
    - EditingZones
    - SoilUse
  - [Owners.xls](#)
  - [TableSoilUse.xls](#)

**Remark:** all data concerning parcels are not real and were created for this academic work on purpose.

## Problem 01

The general goal is to determine the regions vulnerable to the flood risk due to river overflow during the raining season.

Suppose that the vulnerable regions to river flooding are those with height less than or equal to 17.0 m that the main existing river can reach when overflow occurs. The study area extent is defined by the following ETRS PT-TM06 coordinate values:

Xmin= -61800.0 m; Ymin = 37000.0 m; Xmax = -26300.0 m; Ymax = 70000.0 m.

Download the SRTM DEM available at <http://earthexplorer.usgs.gov/> as input data. Consider only the part of the river that gds [River](#) (in geopackage [Territory](#)) represents.

Store all the vector gds created to solve this problem in one file (GeoPackage file format) named [Probl1<student number>](#), under a folder named DataOut. <student number> is the ISA's student

number of one student undersigning the assignment.

Store all the raster gds created to solve this problem in GeoTiff file format under a folder named DataOut.

**1.1.** Create one vector gds representing all vulnerable regions to river flooding in the study area using CRS ETRS-TM06 (EPSG: 3763) and calculate the total area (unit: km<sup>2</sup>) of all these vulnerable regions.

**1.2.** Create one vector gds representing all vulnerable regions from 1. classified as high risk if height is less than 10 m, moderate risk if height is greater than or equal to 10m and less than 15 m, and low risk otherwise. Calculate the total area (unit: ha) of vulnerable regions in each class risk.

**1.3.** Include a diagram representing the integrated solution for Problems 1.1. and 1.2.

## Problem 02

The general goal is to create a new gds representing the parcels, classified by their soil use.

Each group of students will use one (only one) region/polygon of the data set named EditingZones, in the attribute **StudentGroup**, which was assigned by e-mail (review section Context if necessary). The regions are numbered from 1 to 21.

**1.** Create a new gds named **Parcels<n>** (CRS ETRS-TM06 – EPSG: 3763) containing about 10 parcels representing the soil use inside the assigned region based on the Google Satellite image. <n> is the number from 1 to 21 of the assigned region, sent to students by e-mail.

- This gds must be a coverage of the total region (the assigned polygon), i.e. it should cover all the assigned area.
- The allowed soil use codes and corresponding description are in file **TableSoilUse.xls**. Make your best guess to identify the soil use.
- Assign an owner from **Owners.xls** to each parcel (an owner may own more than one parcel but a parcel may only have one owner).
- Store all the vector gds created to solve this problem in one file (GeoPackage file format) named **Probl2<student number>** under a folder named DataOut. <student number> is the ISA's student number of one student undersigning the assignment.

## Problem 03

The general goal is to create a unique gds that includes all parcels in the region delimited by gds **Cadastre** and those created in Problem 02.

The main attributes of these parcels should be soil use (from gds **SoilUse**), and **Parcels<n>** and **owner** (from gds **Cadastre** and **Parcels<n>** created in Problem 02). Concerning soil use, add to the table **TableSoilUse.xls** a new class – **UseCode: 50** and **Description: Other**. Consider all locations from the gds named **Cadastre** not present in the gds named **SoilUse** as soil use code 50 (Other). Store all the vector gds created to solve this problem in one file (GeoPackage file format) named **Probl3<student number>**, under a folder named **DataOut**.

1. Create a new gds (CRS ETRS-TM06 – EPSG: 3763) named **ParcelsNew<student number>** representing the parcels as previously specified. Include a diagram representing the implemented solution.
2. For this new gds build a legend based on soil use.
3. Calculate for each owner his owned total area (units: ha).
4. Calculate for each class of soil use its total area (units: ha).
5. Based on **ParcelsNew<student number>**, create a new gds (CRS ETRS-TM06 – EPSG: 3763) named **Use<student number>** representing only agriculture (*code A*) and forest (*code F*) regions (consider soil use Other neither agriculture nor forest).

## Problem 04

The general goal is to create one gds representing the relative index of susceptibility to groundwater pollution of the region represented in the gds named **GwPollVulnRegion**.

Suppose that this index depends on the soil use and soil type, and its value ( $II_{pp}$ ) is given by the expression

$$II_{pp} = 0.25II_{uu} \times II_{tt}$$

where

$II_{uu}$  – is a factor depending on the soil use; the new gds that you created during Problem 03 represents soil use; Table 1 (see below) contains the values of  $II_{uu}$ ;

$II_{tt}$  – is a factor depending on the soil type; gds named **SoilType** represents soil type; Table 2 (see below) contains the values of  $II_{tt}$ .

Store all the vector gds created to solve this problem in one file (GeoPackage file format) named **Probl4<student number>** under a folder named DataOut.

1. Create the new gds mentioned above (geometry: polygons and CRS ETRS-TM06 – EPSG: 3763) and name it **GwPollIndex**. Show it in the GIS project/map with a legend according to the classes defined in Table 3 (see below). Include a diagram representing the implemented solution.
2. Create a new gds representing locations with  $II_{pp}$  greater than 3 and their owners. Name it **HvHGwPollIndex**.

**Table 1 -  $II_{uu}$  values**

Soil use	$II_{uu}$
Corn	3
Forest	0
Industrial horticulture	6

Soil use	$\Pi_{uu}$
Orchard	1
Other	-1
Vineyard	1
Vegetables	5

Table 2 -  $\Pi_{tt}$  values

Soil type	$\Pi_{tt}$
Modern Luvisol	3
Old Luvisol	1

Table 3 -  $\Pi_{pp}$  values

Index	Class
$< 0$	Unknown
$0 \leq \Pi_{pp} \leq 1$	Very low
$1 < \Pi_{pp} \leq 2$	Low
$2 < \Pi_{pp} \leq 3$	Moderate
$3 < \Pi_{pp} \leq 4$	High
$\Pi_{pp} > 4$	Very high

## Problem 05

In this problem, you will need to be proactive. You will need to identify a problem that should be answered through a spatial analysis. The steps are:

1. Define a question which answer requires the analysis of spatial information. This could be linked to your main area of interest.
2. Collect data from one or several online sources.
3. Clean, organise, merge and process data, as required, so that it can be represented in your GIS project.
4. Normalise you tables according to the 3rd normal form.
5. Map and create a meaningful symbology for the variables you want to analyse.

6. Using your GIS map layout, create a single figure to represent the solution of your problem. If your question has several maps, compose them in the single figure.

*Note: We will learn how to create map layouts in a class hands-on exercise*

7. Include the figure in your report, complemented by a long caption that will answer the question you defined. It should be clear why this problem requires a spatial analysis. Include also the data sources, as citations and references.

For this problem 05, you should create a new GIS project for this problem, independent on the project used for problems 01 to 04.

Examples of questions are:

- Which municipalities in Portugal have the highest areas of cultivated vineyards?
  - data source: Instituto Nacional de Estatística, [www.inec.pt](http://www.inec.pt)
  - You can search for equivalent data in the portal of statistics office of your country
- Which country have the highest production of cereals globally?
  - data source: [FAOSTAT](http://faostat.fao.org/)
- Which cities in Europe have the highest concentration of particles PM10 in the atmosphere, causing health risk?
  - data source: [EEA](http://eea.europa.eu/)

**Notice that the solution of your problem should require *spatial analysis*, and not simply the spatial representation of a factor.**

It is possible that, in order to represent your data on a map, you need to associate it with a spatial layer from a different source. For example, to represent data for cities, you may need to find a map layer with cities, and make data joins between the spatial layer and the data table with the variables of interest.