

**Innovation Program AI4Minerals**  
**Geological and Mining Research Bureau**

Explanatory note - Analysis of geological data for the Ambazac site



## Section 1: Issues and reading the data

### 1. Context and objectives of the analysis

This analysis is part of a data science innovation program applied to geological data, with the main aim of **detecting geochemical anomalies**. These anomalies can be of various kinds (extreme values, atypical spatial behaviour or breaks in expected relationships between elements) and are of great interest from the perspective of exploration, modelling or the identification of areas with particular geological potential.

The study is based on data from the Ambazac area, made available in various formats, including geochemical measurement points, geological structures and layers, and a DTM\* (Digital Terrain Model).

### 2. Presentation of the data set

#### **Data format and sources**

The data can be accessed via WeTransfer on the [program page](#) in the following formats :

1. **Raster (.tif)** for the DTM (Digital Terrain Model)

2. **GeoJSON** for geological structures (L\_GEOL), geological layers (S\_GEOL) and geochemical points
3. Initial processing possible in [QGIS](#), automated reading facilitated by Python libraries (e.g. [rasterio](#), [geopandas](#))

### Description of the main layers

Dataset	Format	Description	Key variables / points of interest
<b>DTM</b>	Raster	Altitude (25 x 25 m pixels)	Direct influence on surface geochemistry
<b>Geological structures (L_GEOL)</b>	GeoJSON	Polylines representing faults and geological contours	Type' field to be filtered to detect relevant objects
<b>Geological layers (S_GEOL)</b>	GeoJSON	Polygons representing geological formations	Use the “Training” field to link it to geochemistry
<b>Geochemical points</b>	GeoJSON	Analysis points with coordinates and ~50 measured elements (with units)	INDC_B, X, Y, concentrations per element (e.g. Au_ppb)

### 3. Initial observations

- **Comprehensive, rich structure:** all the data can be cross-analysed between topography, geological substrate and geochemistry
- **Multi-element geochemical data:** significant potential for studying correlations or unusual profiles (Fe, Cu, Au, etc.)
- **Usable geographical positioning:** each point can be placed in its geological and topographical context
- **Necessary pre-processing to be assessed:** homogeneity of units, treatment of zero values or outliers to be confirmed

*\* A DTM is a computer representation of the ground surface in the form of a grid of points, where each point indicates the altitude at that point. It can be used to visualise and analyse the topography of an area, which is useful for understanding the influence of relief on geochemical data.*

## 4. Next steps

- **Validation of data quality:** consistency of coordinates, verification of units, presence of missing or extreme values, etc.
- **Geospatial cross-referencing:** linking points to surrounding formations and structures
- **Exploratory analysis:** production of descriptive visualisations (histograms, boxplots, interpolated maps) to identify trends or initial anomalies.
- **Preparing to detect anomalies:** according to the three axes defined in the protocol:
  - *amplitude anomalies* : isolated extreme values
  - *spatial anomalies* : atypical values in a local neighbourhood
  - *relational anomalies* : breaks in expected correlations between elements

*\* A DTM is a computer representation of the ground surface in the form of a grid of points, where each point indicates the altitude at that point. It can be used to visualise and analyse the topography of an area, which is useful for understanding the influence of relief on geochemical data.*