

**Concours d'innovation AI4Minerals**  
**Bureau de Recherches Géologiques et Minières**

Notice explicative – Analyse des données géologiques du site d'Ambazac



## Section 1 : Enjeux et lecture des données

### 1. Contexte et objectifs de l'analyse

Cette analyse s'inscrit dans le cadre d'un concours d'innovation data science appliqué à des données géologiques, avec pour objectif principal la **détection d'anomalies géochimiques**. Ces anomalies peuvent être de différentes natures (valeurs extrêmes, comportements spatiaux atypiques ou ruptures dans des relations attendues entre éléments) et sont d'un grand intérêt dans une perspective d'exploration, de modélisation ou d'identification de zones à potentiel géologique particulier.

L'étude porte sur des données issues du secteur d'Ambazac, mises à disposition sous divers formats, incluant notamment des points de mesure géochimiques, des structures et couches géologiques, ainsi qu'un MNT\* (Modèle Numérique de Terrain).

### 2. Présentation du jeu de données

#### **Format et sources des données**

Les données sont accessibles via WeTransfer sur la [page du concours](#) et se présentent sous les formats suivants :

1. **Raster (.tif)** pour le MNT (Modèle Numérique de Terrain)

2. **GeoJSON** pour les structures géologiques (L\_GEOL), les couches géologiques (S\_GEOL) et les points géochimiques
3. Traitement initial possible sous [QGIS](#), lecture automatisée facilitée par des librairies Python (e.g. [rasterio](#), [geopandas](#))

### Description des couches principales

Jeu de données	Format	Description	Variables clés / éléments d'intérêt
MNT	Raster	Altitude (pixels de 25 x 25 m)	Influence directe sur la géochimie de surface
Structures géologiques (L_GEOL)	GeoJSON	Polylignes représentant des failles et contours géologiques	Champ 'Type' à filtrer pour détecter les objets pertinents
Couches géologiques (S_GEOL)	GeoJSON	Polygones représentant des formations géologiques	Champ 'Formation' à exploiter pour relier à la géochimie
Points géochimiques	GeoJSON	Points d'analyse avec coordonnées et ~50 éléments mesurés (avec unités)	INDC_B, X, Y, concentrations par élément (ex. : Au_ppb)

## 3. Premières observations

- **Structure complète et riche** : l'ensemble des données permet une analyse croisée entre topographie, substrat géologique et géochimie
- **Données géochimiques multi-éléments** : potentiel important pour l'étude de corrélations ou de profils inhabituels (Fe, Cu, Au...)
- **Positionnement géographique exploitable** : chaque point peut être replacé dans son contexte géologique et topographique
- **Pré-traitements nécessaires à évaluer** : homogénéité des unités, traitement des valeurs nulles ou aberrantes à confirmer

*\* Un MNT est une représentation informatique de la surface du sol, sous forme de grille de points, où chaque point indique l'altitude à cet endroit. Il permet de visualiser et analyser la topographie d'un territoire, ce qui est utile pour comprendre l'influence du relief sur les données géochimiques*

## 4. Prochaines étapes envisagées

- **Validation de la qualité des données** : cohérence des coordonnées, vérification des unités, présence de valeurs manquantes ou extrêmes
- **Croisement géospatial** : rattachement des points aux formations et structures environnantes
- **Analyse exploratoire** : production de visualisations descriptives (histogrammes, boxplots, cartes interpolées) pour dégager des tendances ou premières anomalies
- **Préparation à la détection d'anomalies** : selon trois axes définis dans le protocole :
  - *anomalies d'amplitude* : valeurs extrêmes isolées
  - *anomalies spatiales* : valeurs atypiques dans un voisinage local
  - *anomalies relationnelles* : ruptures dans des corrélations attendues entre éléments

\* Un MNT est une représentation informatique de la surface du sol, sous forme de grille de points, où chaque point indique l'altitude à cet endroit. Il permet de visualiser et analyser la topographie d'un territoire, ce qui est utile pour comprendre l'influence du relief sur les données géochimiques