

## **Génération de nouvelles visualisations via la fonction UMAP (méthode non-linéaire)**

### **Les deux paramètres principaux d'UMAP**

#### **1. n\_neighbors**

- **Rôle :**  
Ce paramètre détermine la taille du voisinage local utilisé pour construire l'ensemble fuzzy simplicial. Autrement dit, il indique combien de voisins autour de chaque point sont pris en compte pour capturer la structure locale des données.
- **Valeurs recommandées :**  
Typiquement, on fait varier ce paramètre entre 5 et 50. Une valeur par défaut courante est 15.
- **Effets :**
  - **Valeurs faibles (par exemple, 5) :**  
Elles favorisent la préservation des détails locaux et peuvent révéler des sous-structures fines. Toutefois, cela peut aussi rendre l'embedding plus « bruité » sur l'échelle globale.
  - **Valeurs élevées (par exemple, 50) :**  
Elles mettent l'accent sur la structure globale, ce qui lisse les détails locaux et donne une représentation plus agrégée des données.

#### **2. min\_dist**

- **Rôle :**  
Ce paramètre définit la distance minimale entre les points dans l'espace réduit. Il contrôle la compacité des groupes de points dans l'embedding en faible dimension.
- **Valeurs recommandées :**  
Les valeurs varient généralement de 0.0 à environ 0.5. Par défaut, on trouve souvent des valeurs comprises entre 0.1 et 0.3.
- **Effets :**
  - **Valeur proche de 0 (ex. 0.0–0.1) :**  
Permet aux points d'être très proches les uns des autres, révélant ainsi des zones de forte densité locale.
  - **Valeur plus élevée (ex. 0.3–0.5) :**  
Force les points à être plus espacés, ce qui peut rendre l'embedding plus dispersé et accentuer la séparation entre les différents clusters.

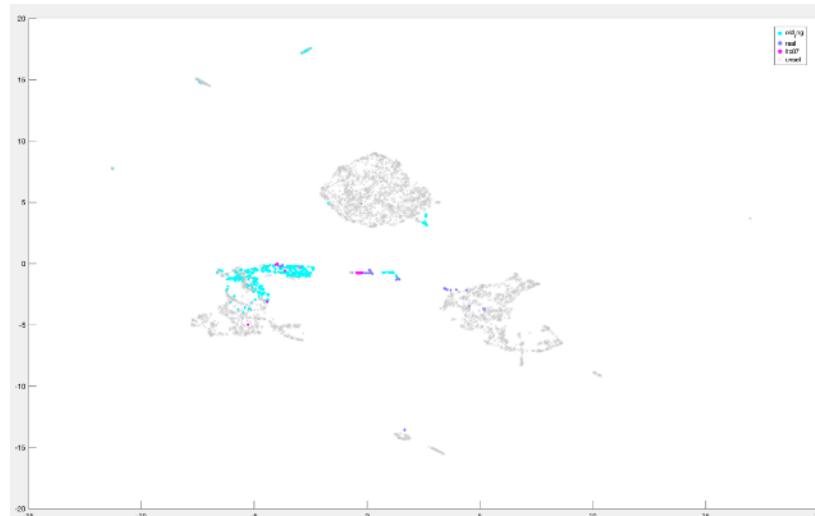
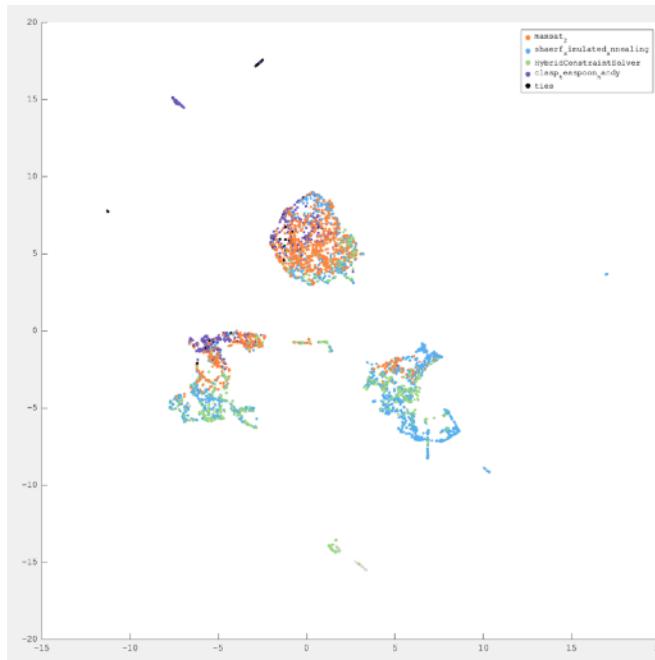
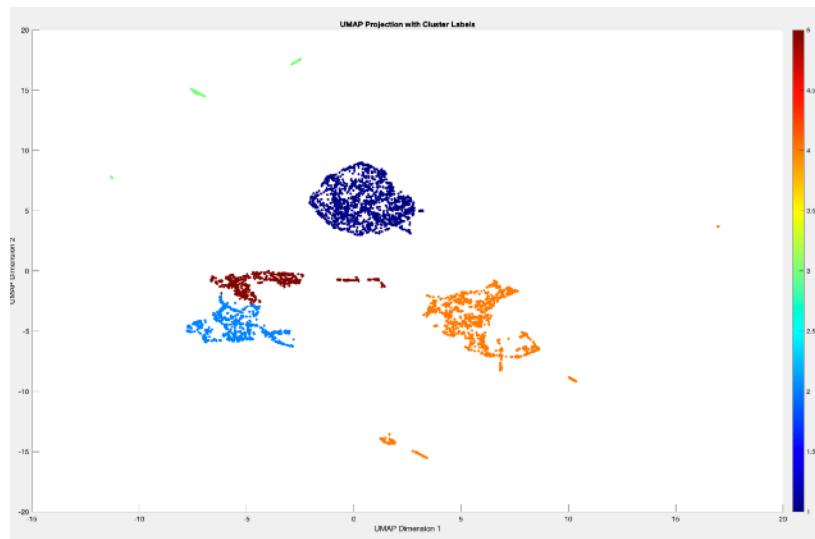
#### **Résumé :**

Pour explorer différentes visualisations avec UMAP, on peut expérimenter en faisant varier :

- **n\_neighbors** entre 5 et 50, qui influence l'équilibre entre la structure locale et globale des données.
- **min\_dist** entre 0.0 et 0.5, qui détermine la compacité des clusters dans l'espace réduit.

L'ajustement de ces paramètres doit se faire en fonction du niveau de détail ou de globalité que l'on souhaite observer dans l'embedding.

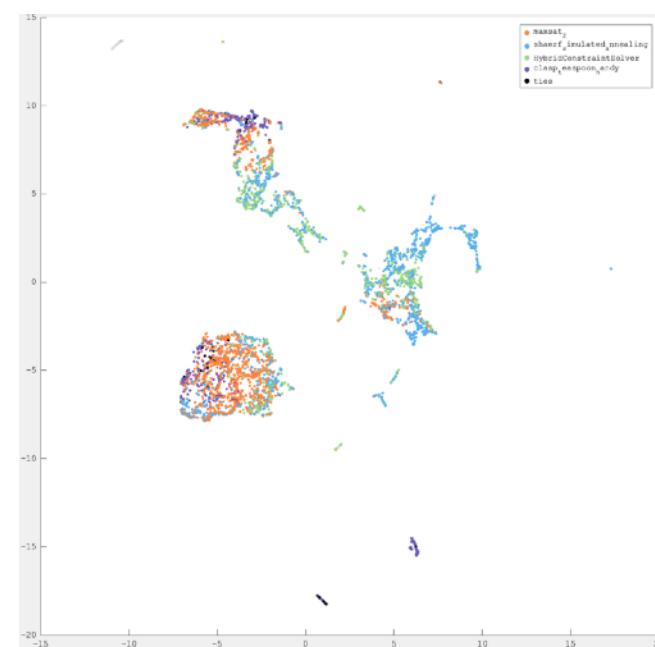
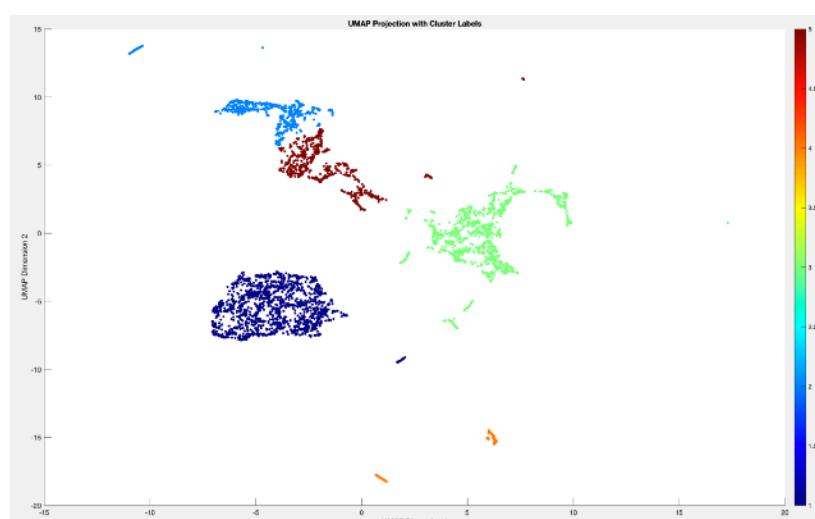
*n\_neighbors = 15 et min\_dist = 0.1*

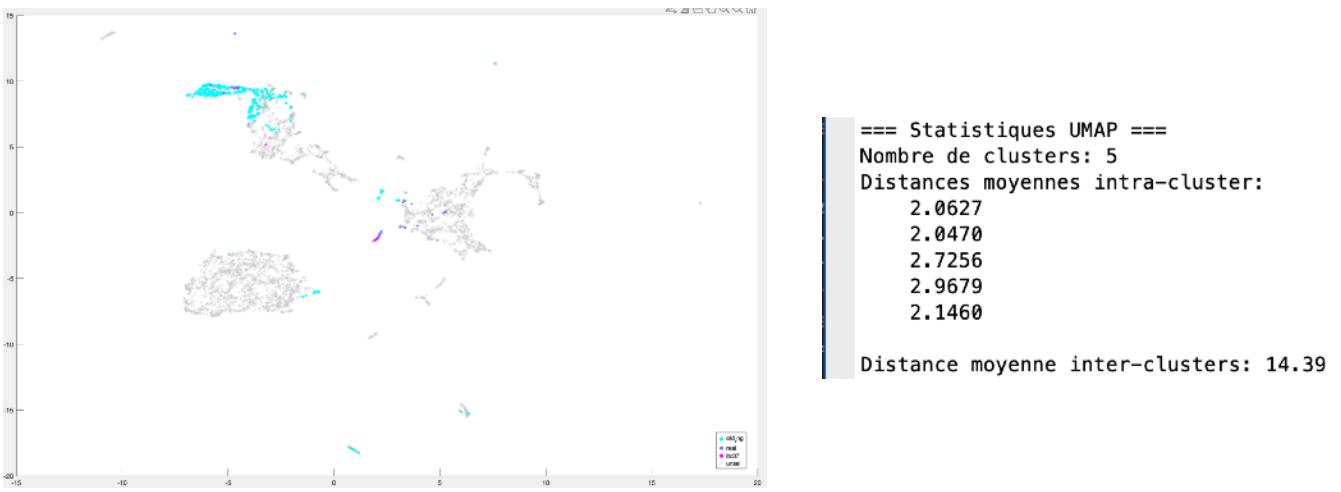


```
==== Statistiques UMAP ====
Nombre de clusters: 5
Distances moyennes intra-cluster:
 1.8267
 1.4581
 3.2739
 3.2524
 1.7616

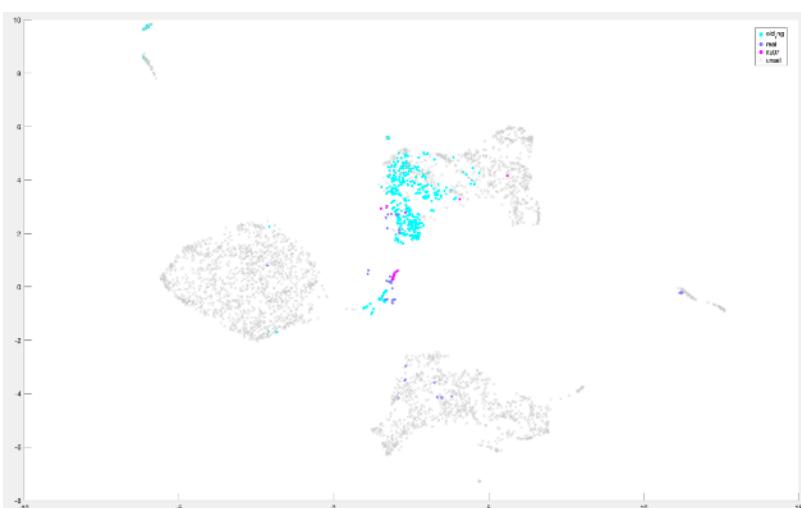
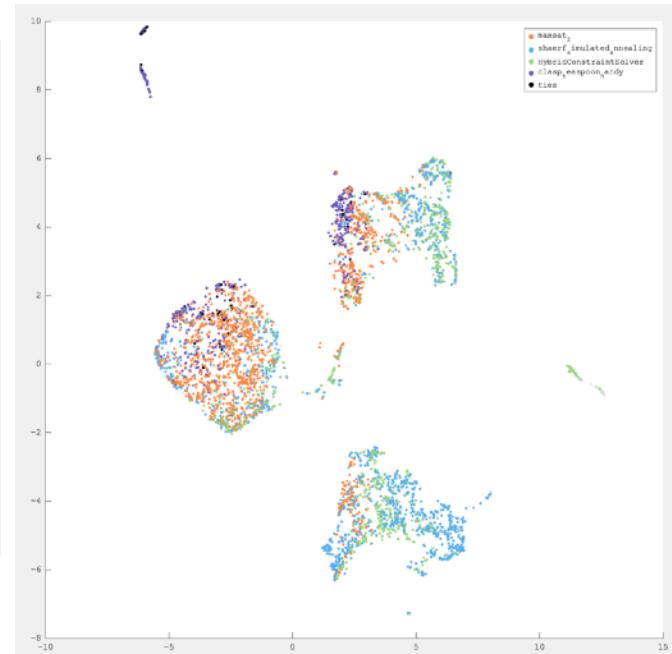
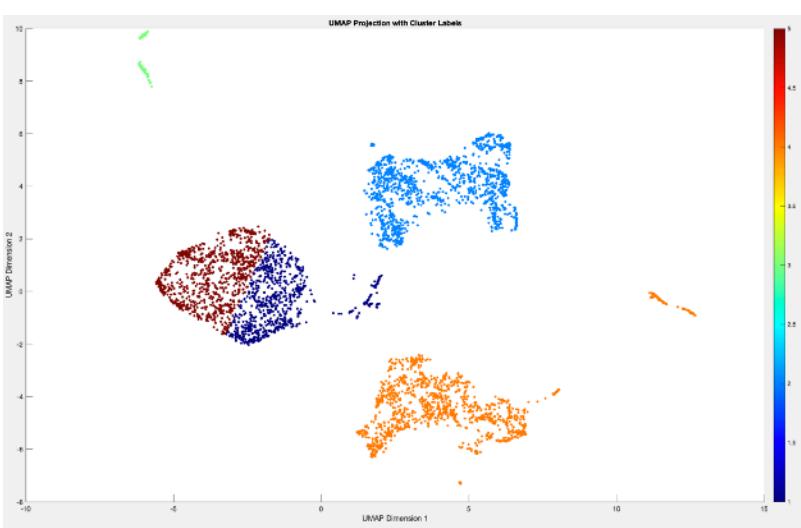
Distance moyenne inter-clusters: 12.82
```

*n\_neighbors = 10 et min\_dist = 0.1*

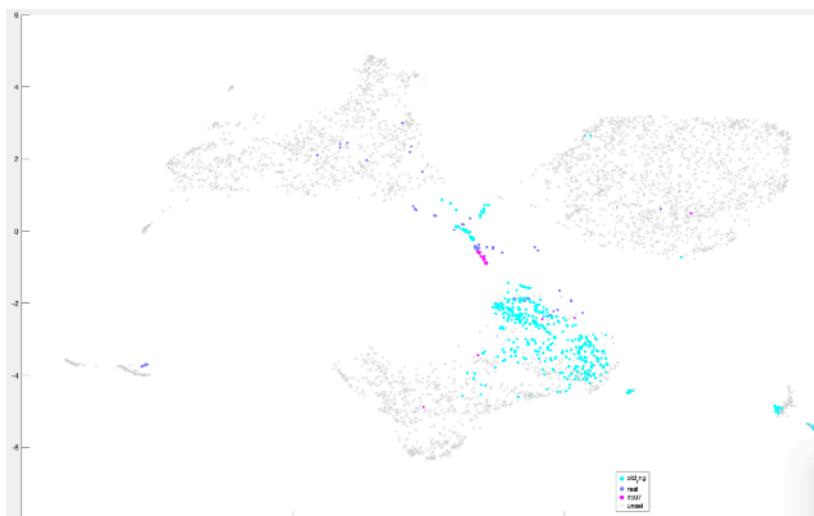
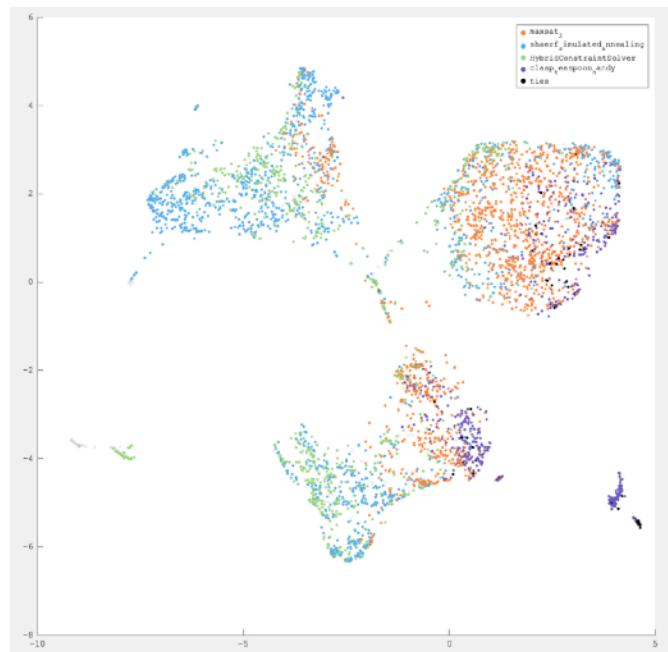
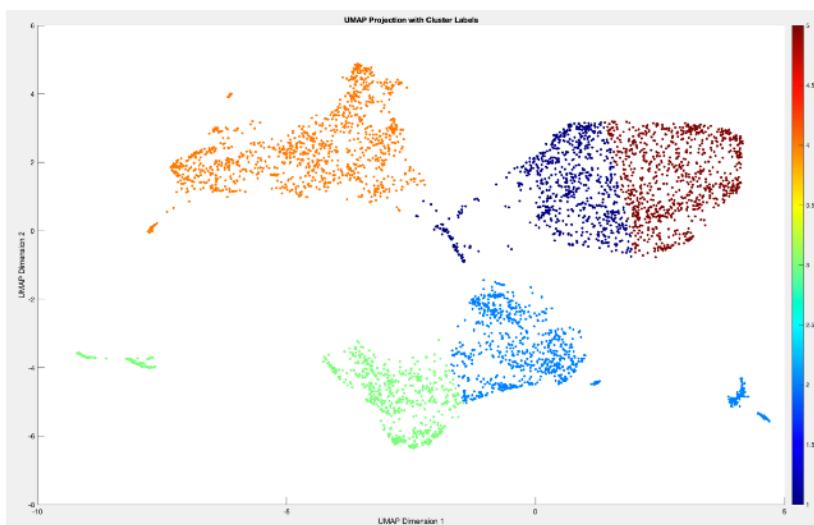




$n\_neighbors = 50$  et  $min\_dist = 0.1$

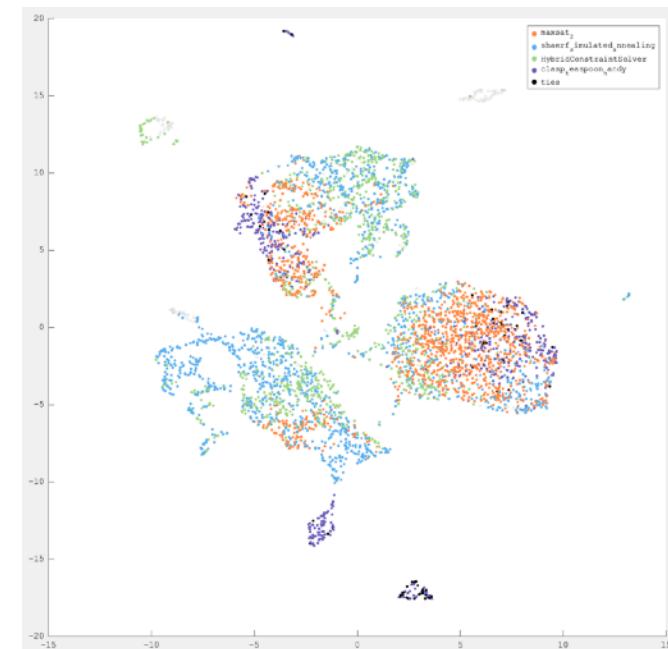
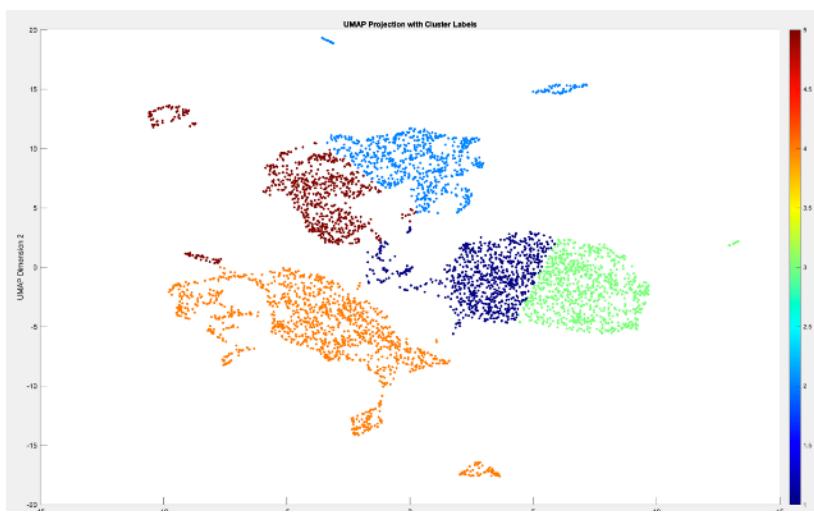


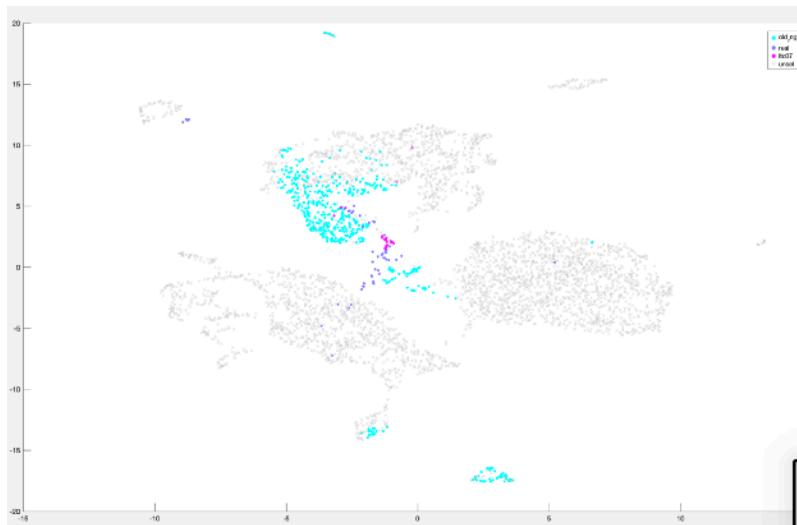
*n\_neighbors = 100 et min\_dist = 0.1*



```
==== Statistiques UMAP ====
Nombre de clusters: 5
Distances moyennes intra-cluster:
 1.2777
 1.6938
 1.9306
 1.6310
 1.2061
Distance moyenne inter-clusters: 6.17
```

*n\_neighbors = 10 et min\_dist = 0.5*





==== Statistiques UMAP ===

Nombre de clusters: 5

Distances moyennes intra-cluster:

2.3000

2.8353

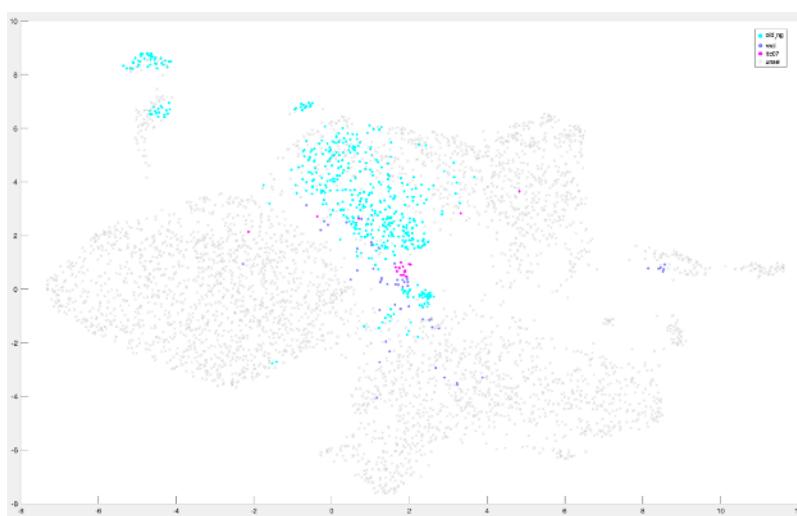
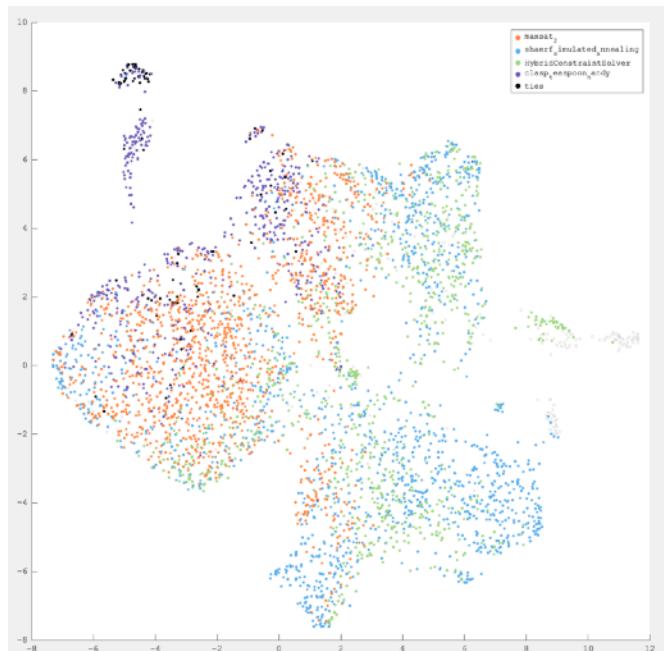
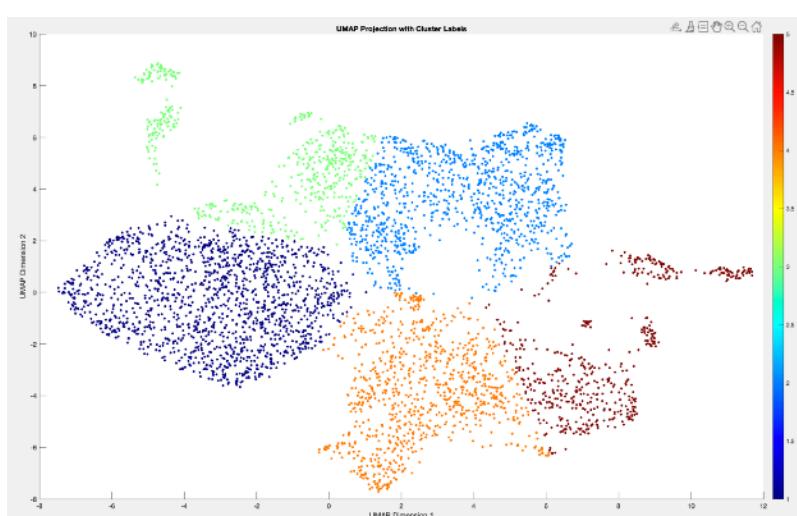
2.2878

4.0589

3.1641

Distance moyenne inter-clusters: 10.53

*n\_neighbors = 50 et min\_dist = 0.5*



==== Statistiques UMAP ===

Nombre de clusters: 5

Distances moyennes intra-cluster:

2.2837

2.2144

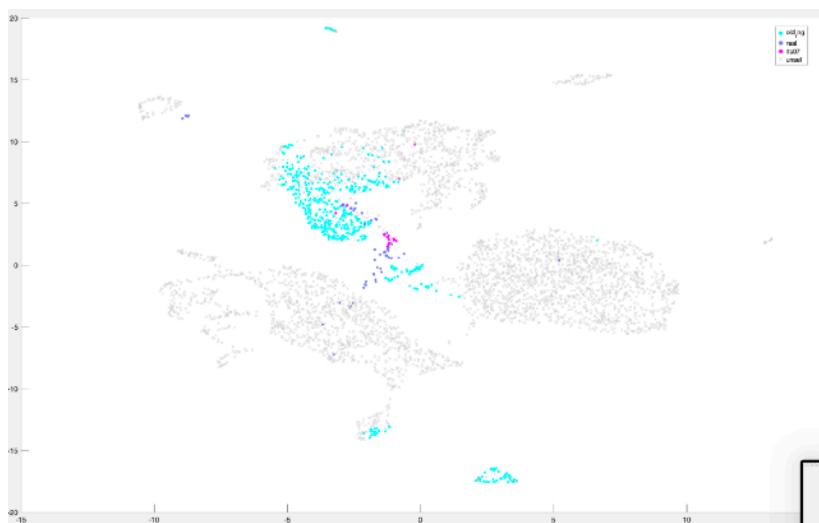
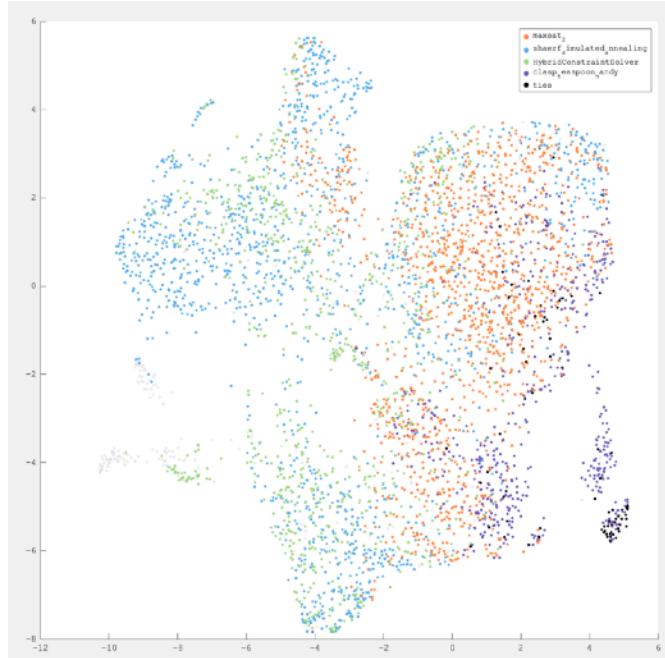
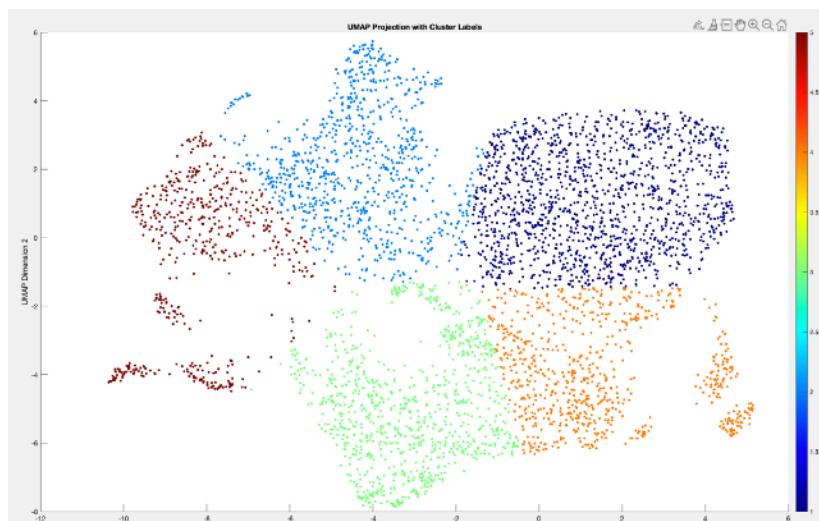
2.5434

2.1315

2.5827

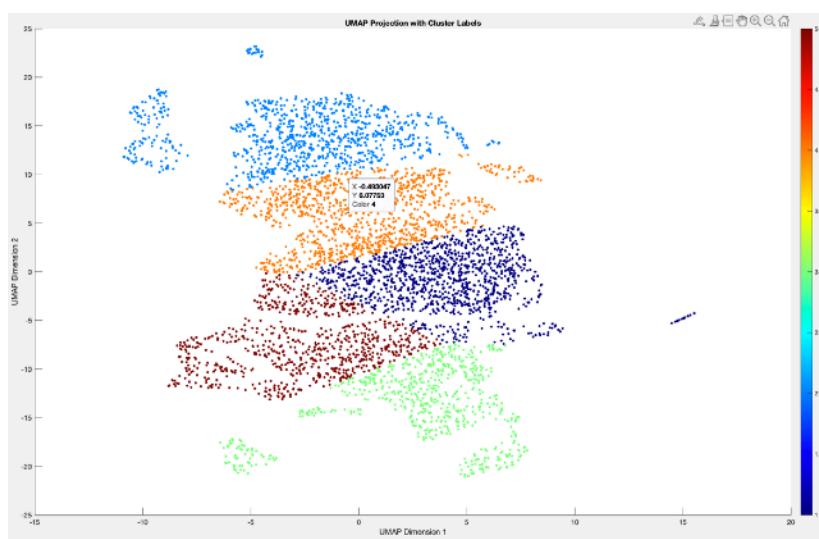
Distance moyenne inter-clusters: 7.87

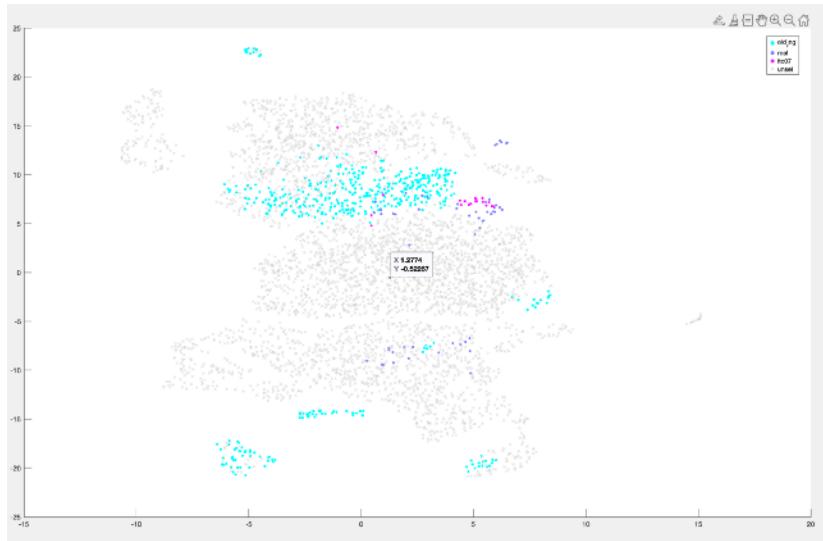
***n\_neighbors = 100 et min\_dist = 0.5***



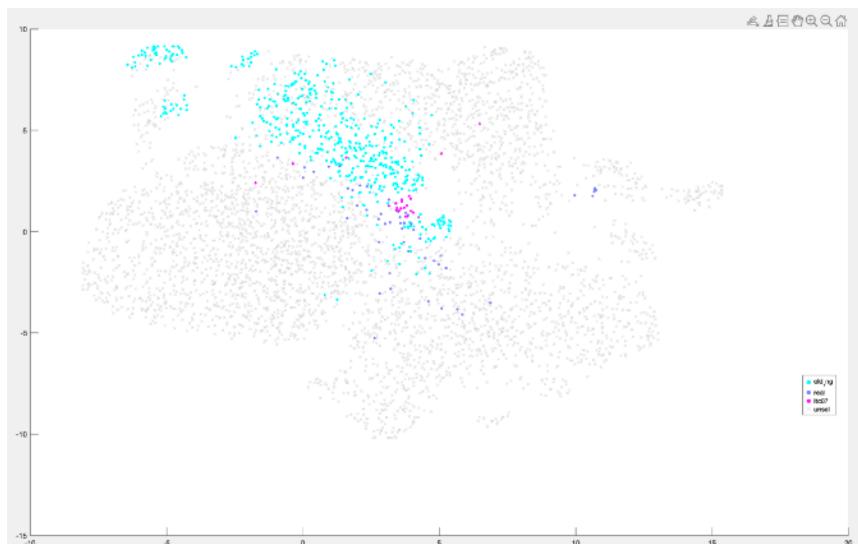
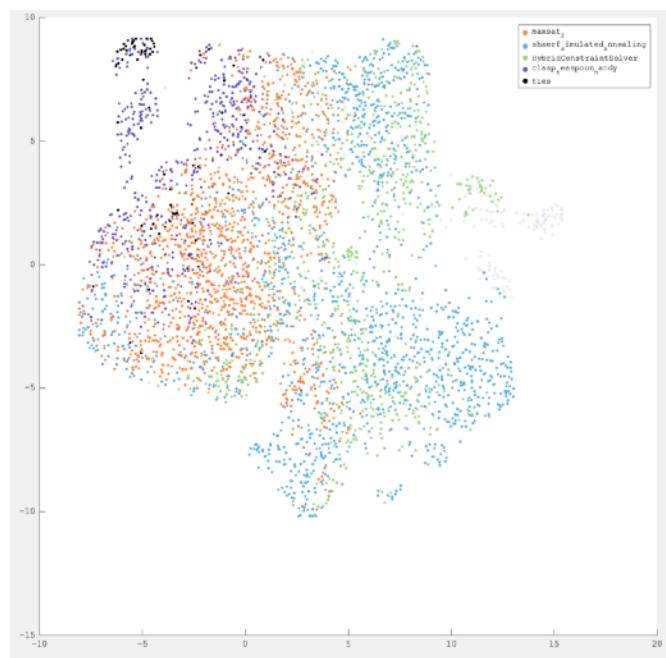
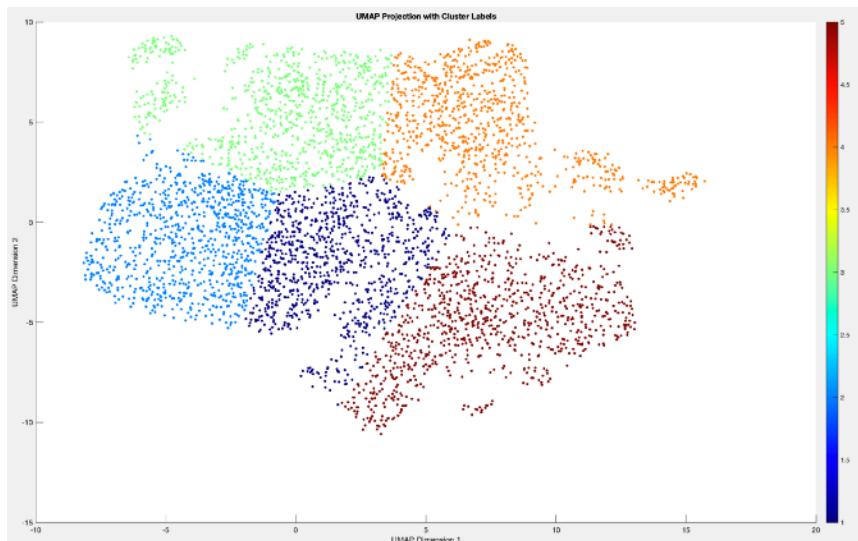
==== Statistiques UMAP ====  
**Nombre de clusters:** 5  
**Distances moyennes intra-cluster:**  
 2.0733  
 1.9573  
 1.9865  
 2.0406  
 2.2371  
**Distance moyenne inter-clusters:** 6.96

***n\_neighbors = 10 et min\_dist = 1***

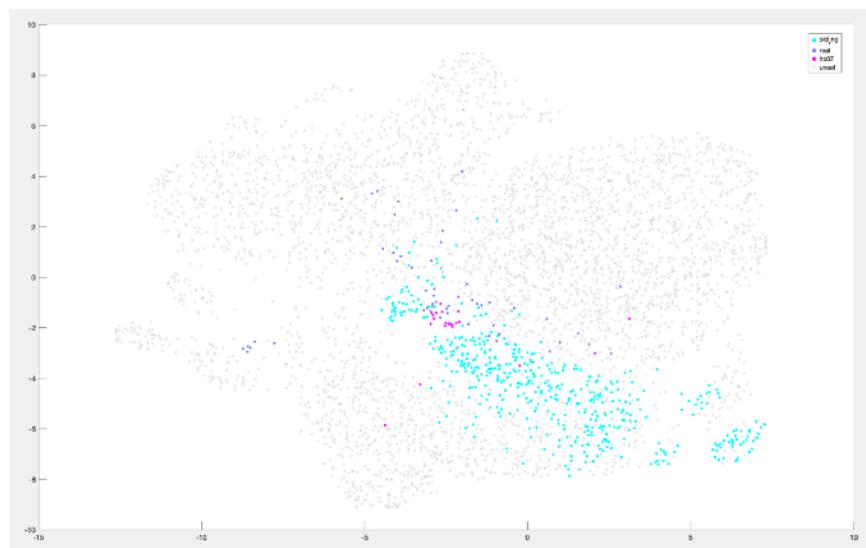
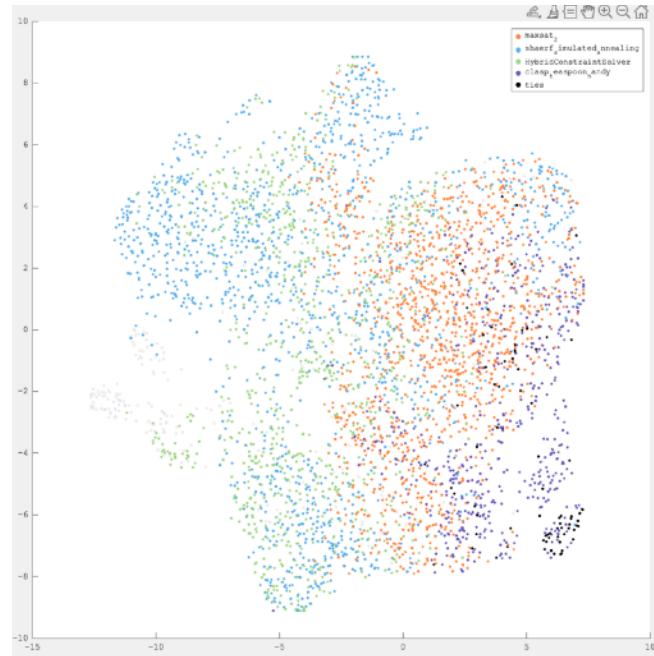
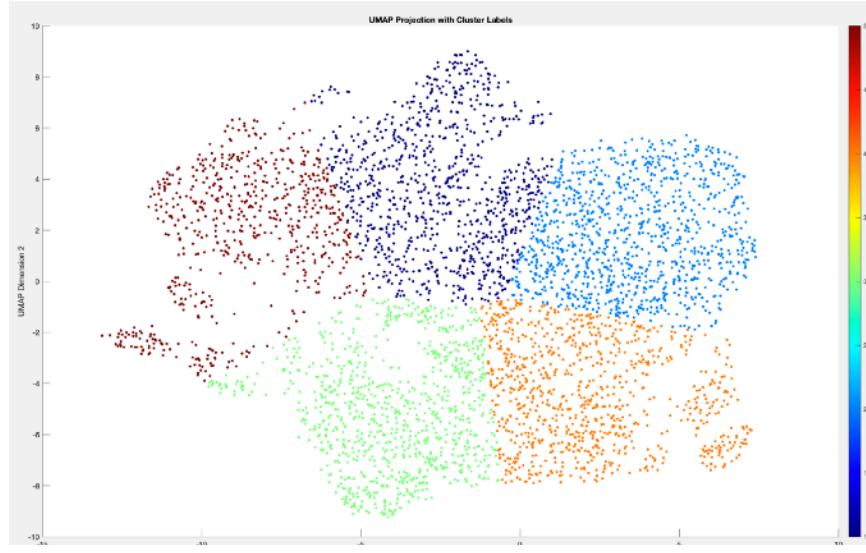




*n\_neighbors = 50 et min\_dist = 1*



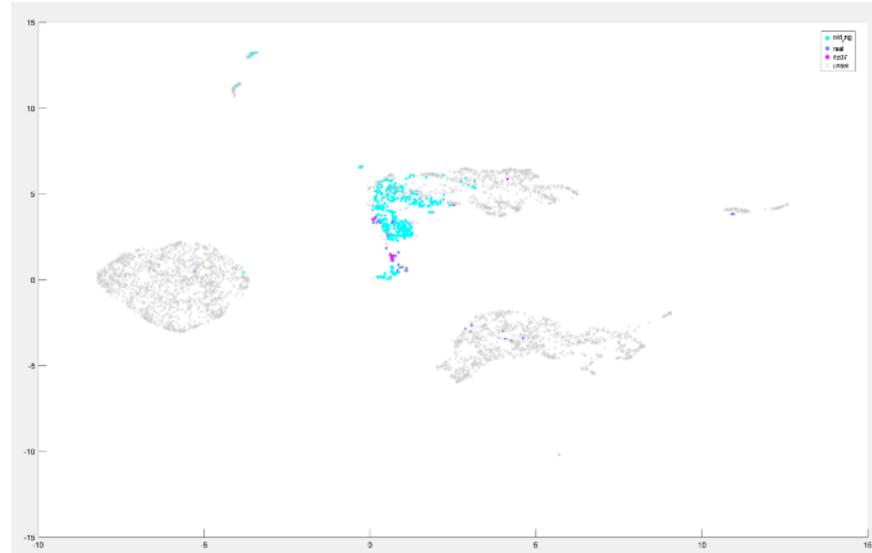
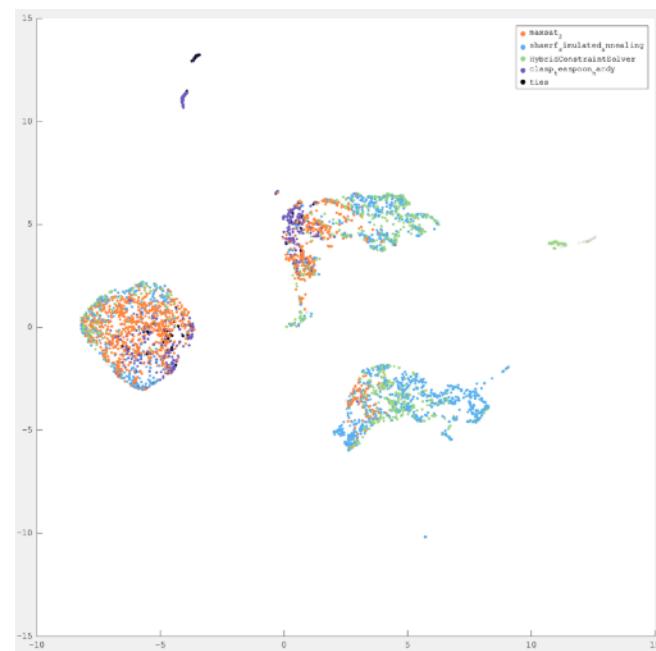
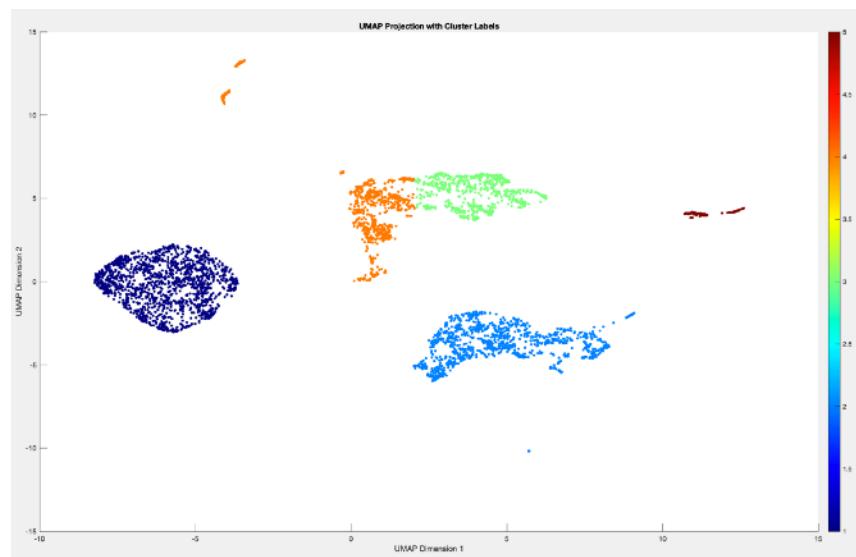
***n\_neighbors = 100 et min\_dist = 1***



==== Statistiques UMAP ====  
Nombre de clusters: 5  
Distances moyennes intra-cluster:  
2.7837  
2.5379  
2.6574  
2.7273  
2.8736

Distance moyenne inter-clusters: 8.63

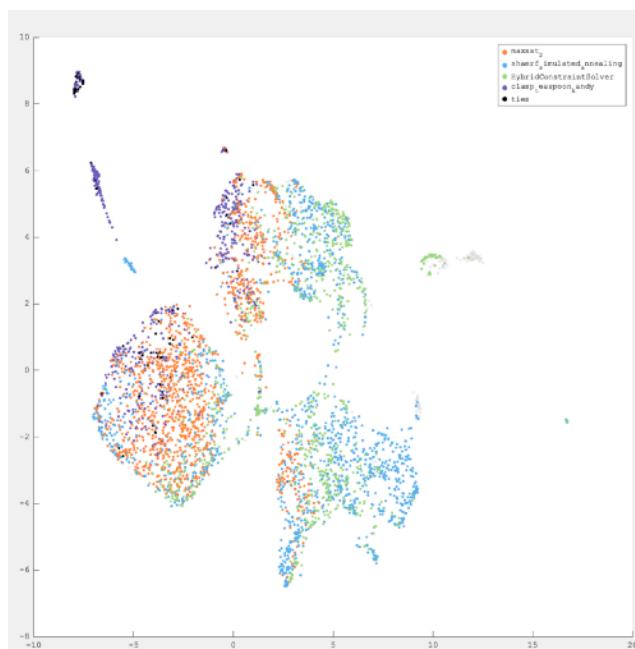
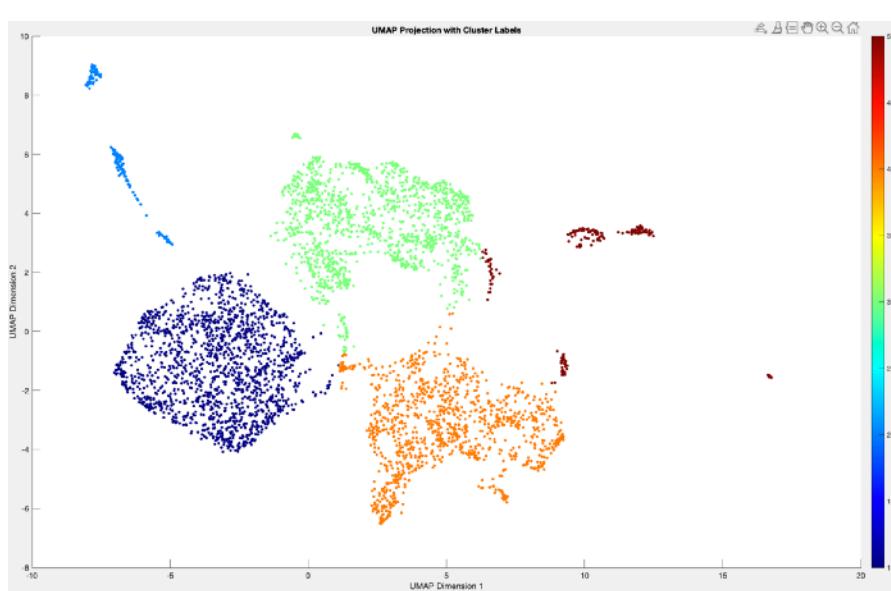
*n\_neighbors = 30 et min\_dist = 0.1*

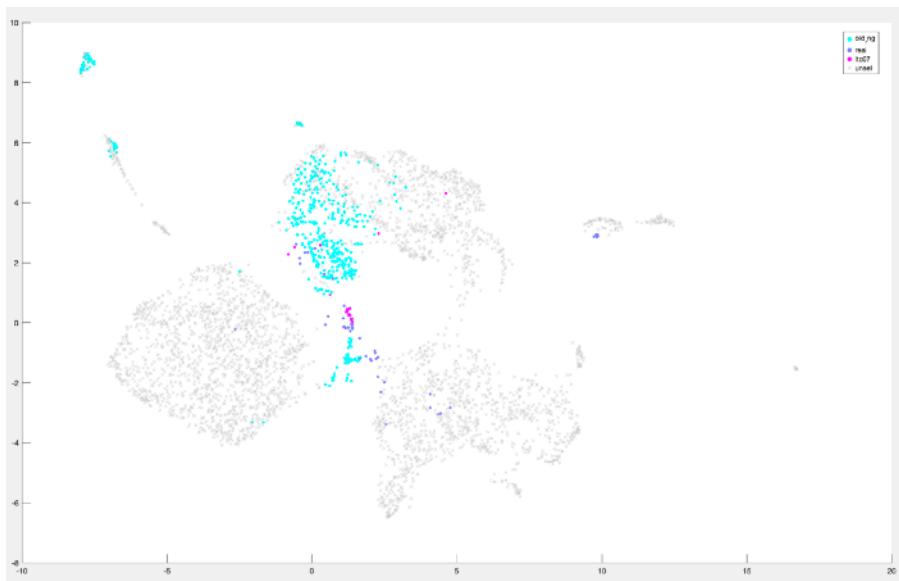


== Statistiques UMAP ==  
 Nombre de clusters: 5  
 Distances moyennes intra-cluster:  
 1.6378  
 1.9302  
 1.2239  
 2.9273  
 0.6450

Distance moyenne inter-clusters: 10.22

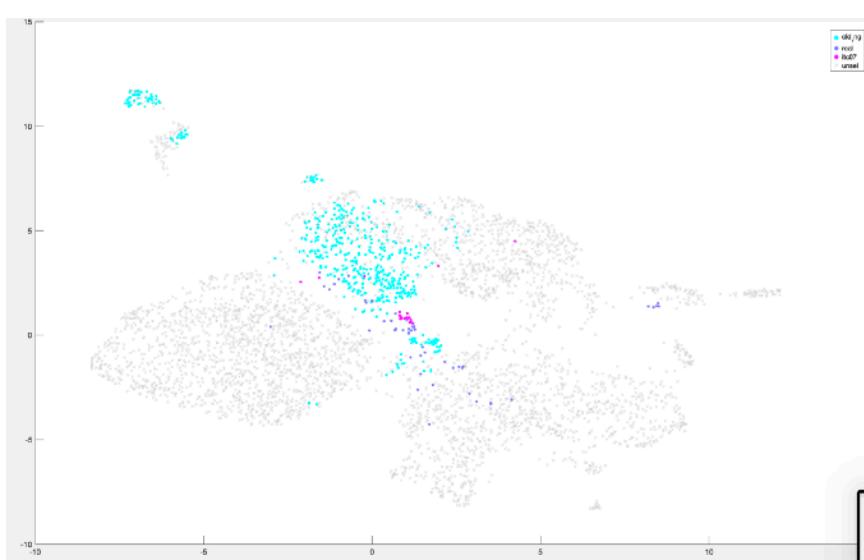
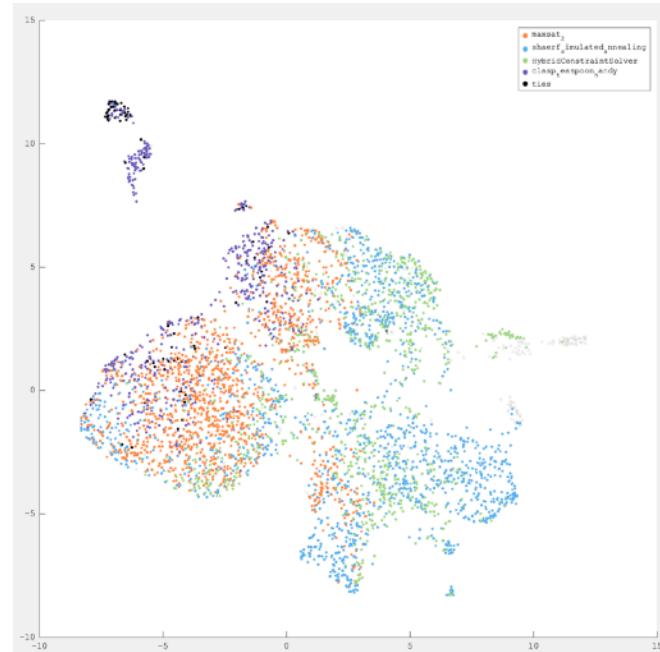
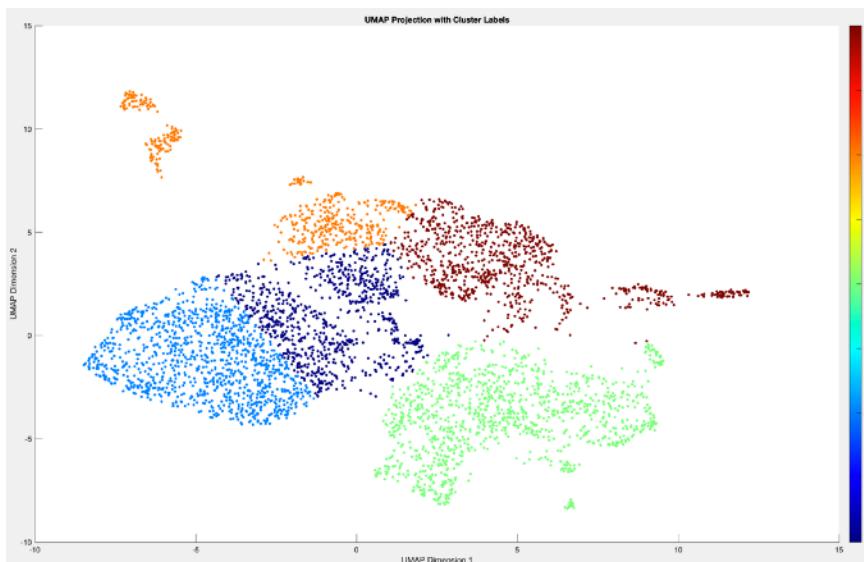
*n\_neighbors = 30 et min\_dist = 0.3*





**Nombre de clusters: 5**  
**Distances moyennes intra-cluster:**  
 2.0637  
 1.8912  
 2.2759  
 2.2458  
 2.7298  
**Distance moyenne inter-clusters: 10.31**

**$n\_neighbors = 30$  et  $\text{min\_dist} = 0.5$**



**== Statistiques UMAP ==**  
**Nombre de clusters: 5**  
**Distances moyennes intra-cluster:**  
 2.1929  
 2.2044  
 2.7223  
 3.2300  
 2.4988

**Distance moyenne inter-clusters: 8.07**