

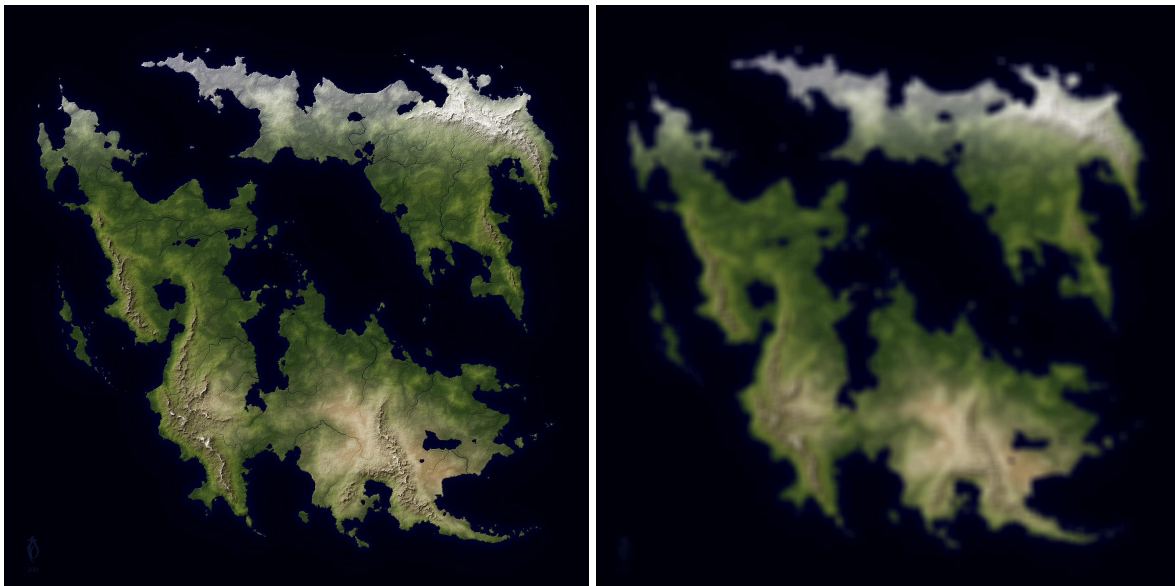
Rapport SAE - Détection de biomes sur des exoplanètes

Un rapport concis mais complet avec la présentation du travail effectué et une analyse des résultats obtenus

I - Les flous

On a commencé par créer 2 algorithmes de floutage, un par moyenne et un gaussien. Ces flous nous ont servi à rendre nos algorithmes de clustering beaucoup plus efficaces en enlevant les détails des images.

Voici la différence entre l'image d'origine et l'image floutée :



II - Les algorithmes de clustering

Pour réaliser ce travail, nous nous sommes répartis un algorithme par personne. Ces algorithmes de clustering nous ont permis de regrouper les pixels en fonction de la similarité de leur couleur. Ce qui nous a aussi permis de distinguer les différents biomes d'une planète grâce à la palette de couleurs fournis.

1. Choix de l'algorithme de clustering le plus adapté à cette tâche. Prenez en compte notamment le nombre de pixels à traiter.

Après avoir testé les algorithmes suivants :

- K-Means
- DBSCAN
- HAC

Nous avons remarqué que K-Means était le plus adapté pour cette tâche car DBSCAN et HAC sont beaucoup trop longs pour traiter de grandes images. Pour une image de 100x100 K-Means peut la clusteriser en 1 seconde et DBScan en 3 minutes.

2. Choix de la métrique pour calculer la similarité entre couleurs (voir TP de préparation).

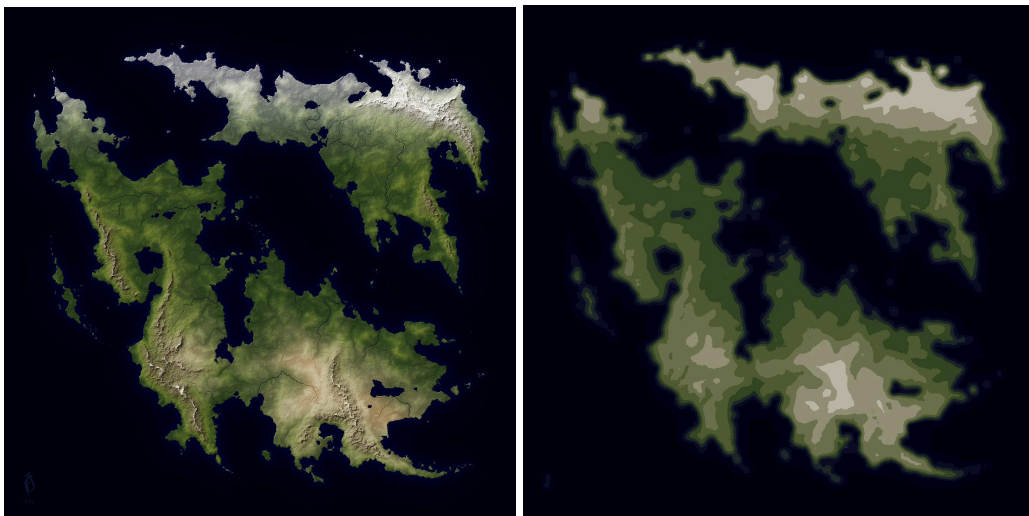
On a utilisé la norme normale.

3. Choix des paramètres du clustering. Il est conseillé de tester différentes valeurs.

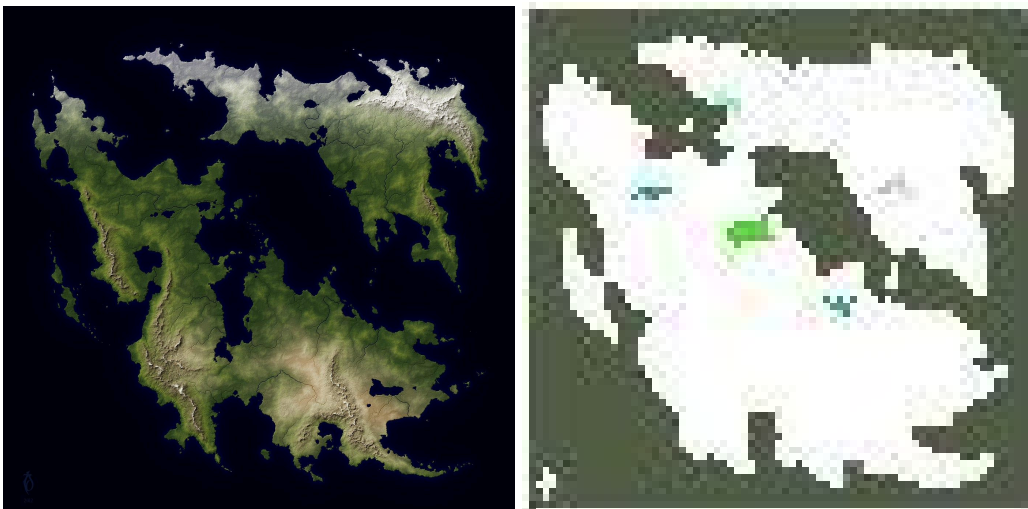
On a choisi de détailler les images en 10 clusters car on a eu au préalable une palette de 10 couleurs. On a donc chaque cluster qui est assigné à un biome.

4. Évaluation de la qualité du résultat obtenu. La validation peut être visuelle ou en utilisant des indices de qualité.

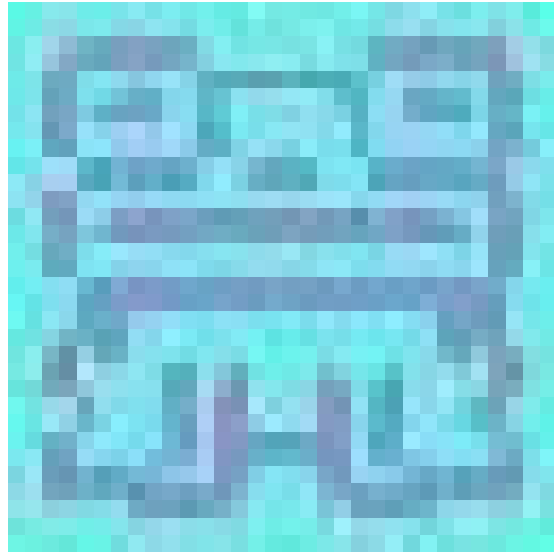
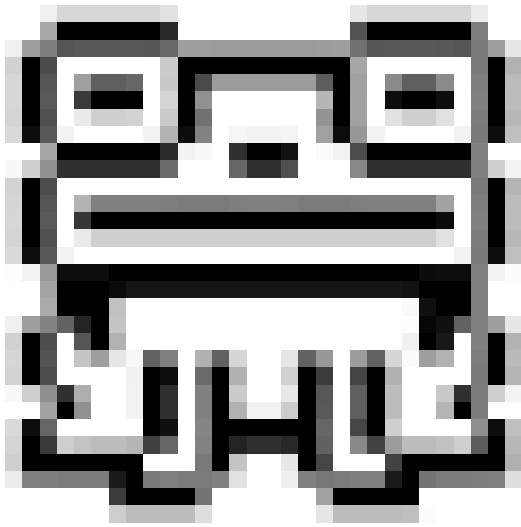
K-Means



DBSCAN



HAC



III - Pourquoi on a choisi KMeans

KMeans est un algorithme de clustering basé sur la minimisation de la variance intra-cluster. Il divise les données en K clusters en initialisant des centroids, puis en itérant des phases d'assignation et de mise à jour des centroids jusqu'à convergence.

Les avantages :

- Simple à implémenter.
- Efficace et rapide pour des grands ensembles de données.
- Les résultats sont facilement interprétables.

Inconvénients :

- Nécessite de spécifier le nombre de clusters à l'avance.
- Sensible aux choix initiaux des centroids.
- Reconnaît des clusters de forme sphérique.