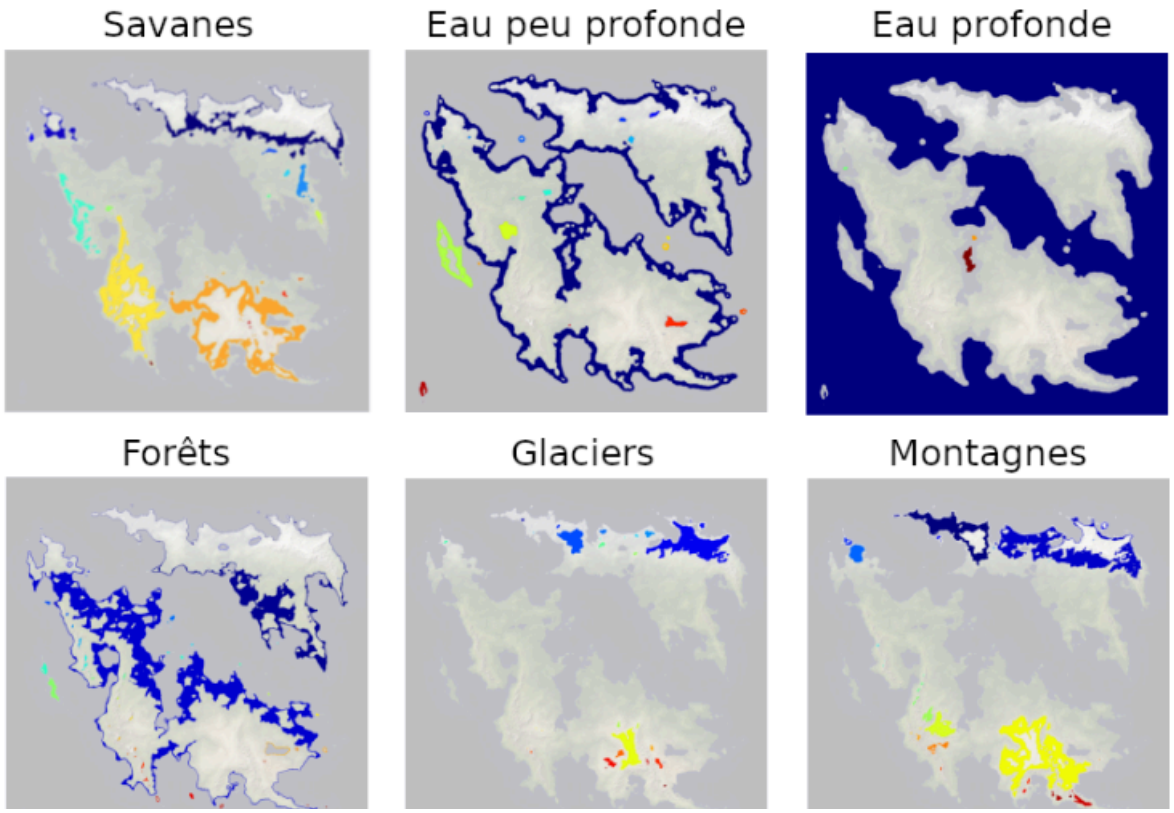


Détection de biomes sur des exoplanètes

20/06/2024

KOMODZINSKI Jawad
LOPPINET Stéphane
RYSAK Hugo
TROHA Stanislas

Objectifs

1. Détecter automatiquement les différents biomes présents sur une exoplanète
2. Définir les différents écosystèmes appartenant à chaque biome.

Grandes étapes

I. Prétraitement de l'image

Lisser et d'homogénéiser l'image afin de faire disparaître les petits détails qui ne nous intéressent pas.

II. Détection et visualisation des biomes

Utiliser des algorithmes de clustering pour regrouper les pixels en fonction de la similarité de leur couleur.

III. Détection et affichage des écosystèmes pour chaque biome

Utiliser un algorithme de clustering sur la position (abscisses et ordonnées dans l'image) des pixels de chaque biome avec une distance euclidienne.

1. Prétraitement de l’image

Tout d'abord, le flou gaussien est optimal, car il permet de réduire les détails indésirables tout en préservant les contours essentiels de l'image. Cela est particulièrement important pour garantir que les limites entre les différents biomes et écosystèmes restent clairement définies et facilement identifiables par les algorithmes de clustering.
De plus, le flou gaussien est bien adapté à la perception humaine du flou, ce qui signifie que les résultats seront plus intuitifs et plus faciles à interpréter visuellement.

2. Détection et visualisation des biomes

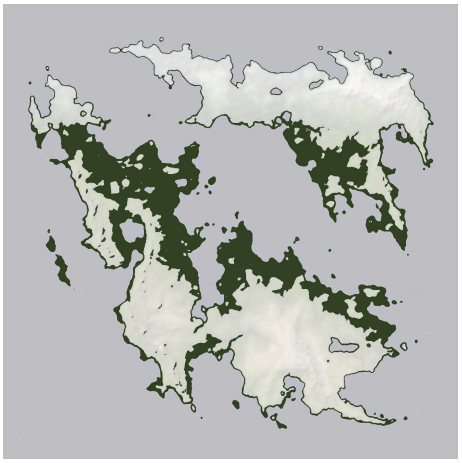
Pour notre projet de détection des biomes, nous avons choisi d'utiliser l'algorithme K-Means en raison de sa complexité moindre et de sa simplicité de mise en place, ce qui en fait une première étape pratique et efficace. Cependant, nous reconnaissons que l'algorithme DBSCAN pourrait s'avérer particulièrement intéressant pour mieux gérer les formes irrégulières des biomes naturels.
Aussi, l'algorithme HAC offre des avantages supplémentaires en permettant non seulement de traiter les formes irrégulières, mais aussi de créer une hiérarchie des clusters, ce qui facilite l'identification des sous-catégories de biomes, comme les forêts tropicales et tempérées. Néanmoins, le temps d'exécution de DBSCAN est déjà immense (complexité $O(n^2)$), alors passer sur HAC avec une complexité de $O(n^3)$ semble un peu présomptueux.

Pour ce qui est des normes, nous avons choisi de commencer par utiliser la distance euclidienne dans l'espace RGB pour calculer les similarités de couleur, principalement en raison de sa simplicité et de sa faible complexité computationnelle. Cependant, nous reconnaissons que pour améliorer la qualité des nuances et obtenir une détection de biomes plus précise, les normes Redmean et CIELAB pourraient s'avérer plus pertinentes. Redmean offre un compromis intéressant en tenant partiellement compte de la perception humaine des couleurs sans ajouter une complexité excessive, tandis que CIELAB, bien que plus complexe à mettre en œuvre, pourrait fournir la meilleure différenciation des couleurs en raison de son uniformité perceptuelle.

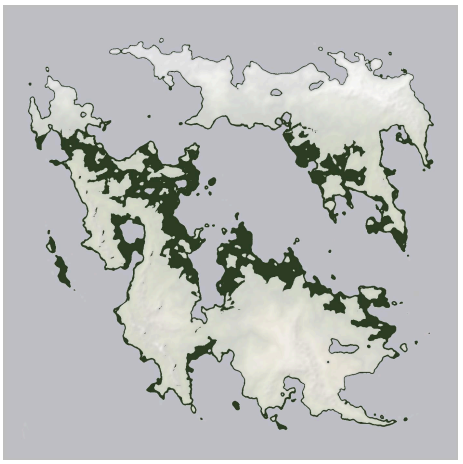
Voici les résultats obtenus avec différentes normes et sur différentes planètes. Nous montrerons les biomes “forêt” et “océan”.

-Planète 1 :

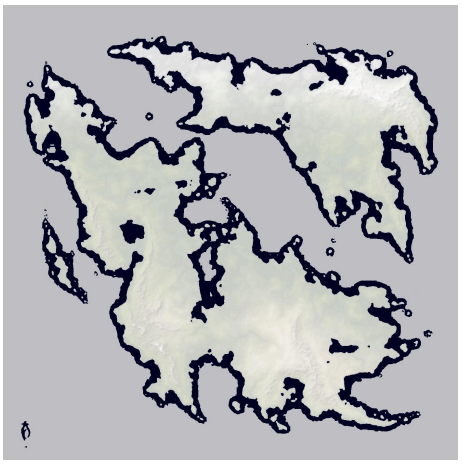
-Normal :

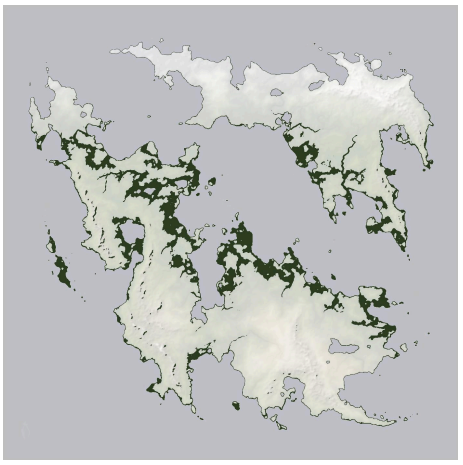


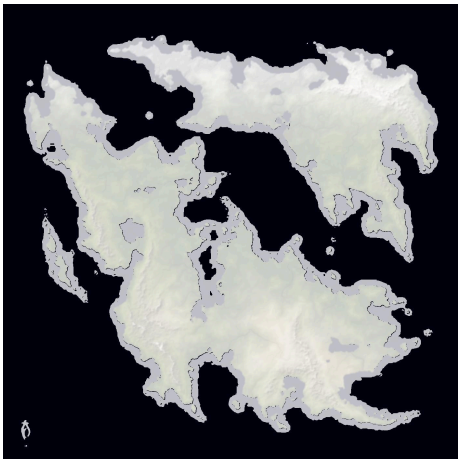
-RedMean :



-CieLab:



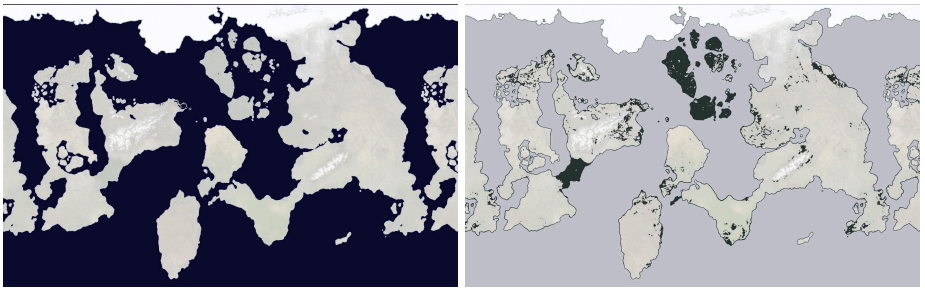




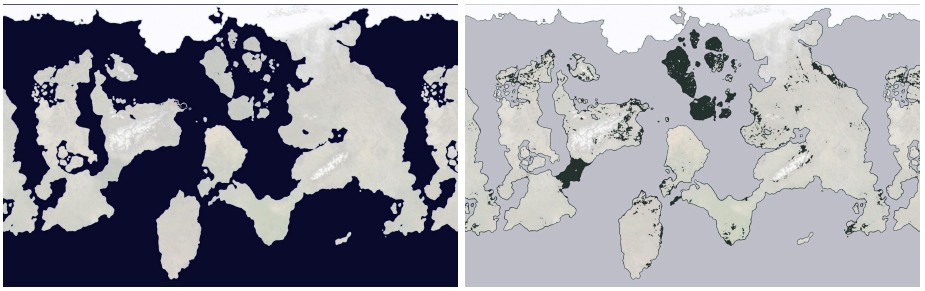
On peut voir que ici CieLab fait bien le contraste entre eau profondes et eau peu profondes et que le biome forêt est plus petit, cela veut dire que RedMean et RGB Basique font des différences sur d'autre biomes tels que les biome aride ou rocheux ou la différence se voit plus que dans CieLab

- Planète 2 :

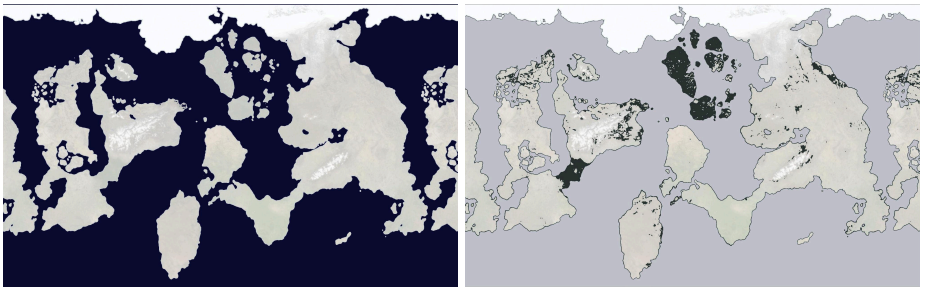
-Normal :



-RedMean :

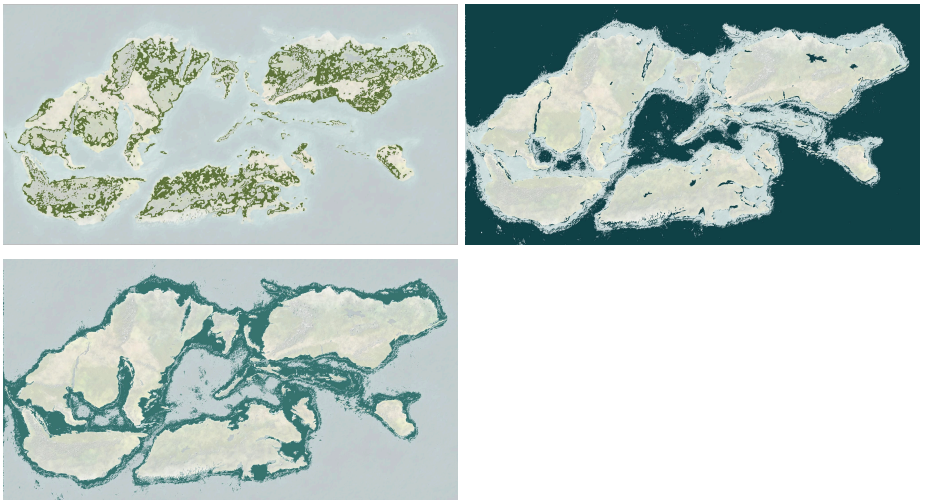


-CieLab :

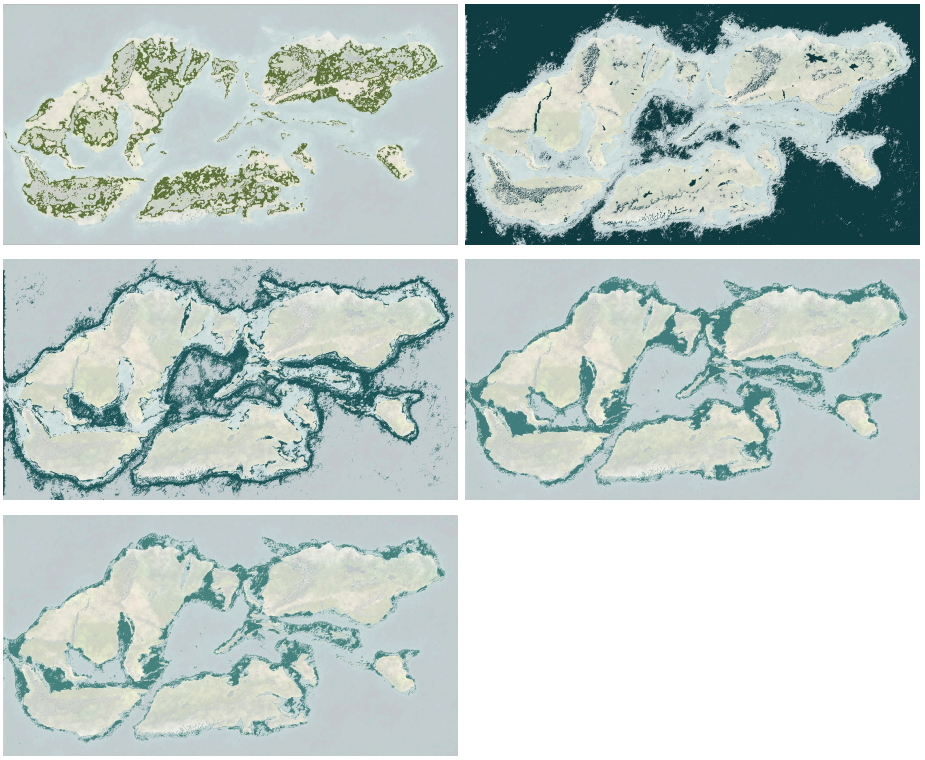


-Planete 3 :

-Normal :

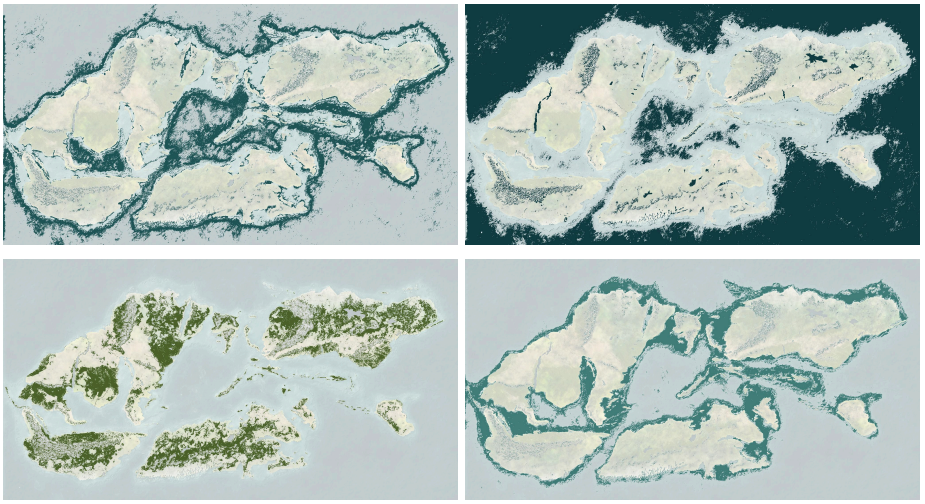


-RedMean :



(Les différences se voient plutôt sur les océans où Redmean semble relever plus de contraste de couleurs que celui de base)

-CieLab :



Les différentes normes ne semblent pas toujours relever des différences dans les océans.

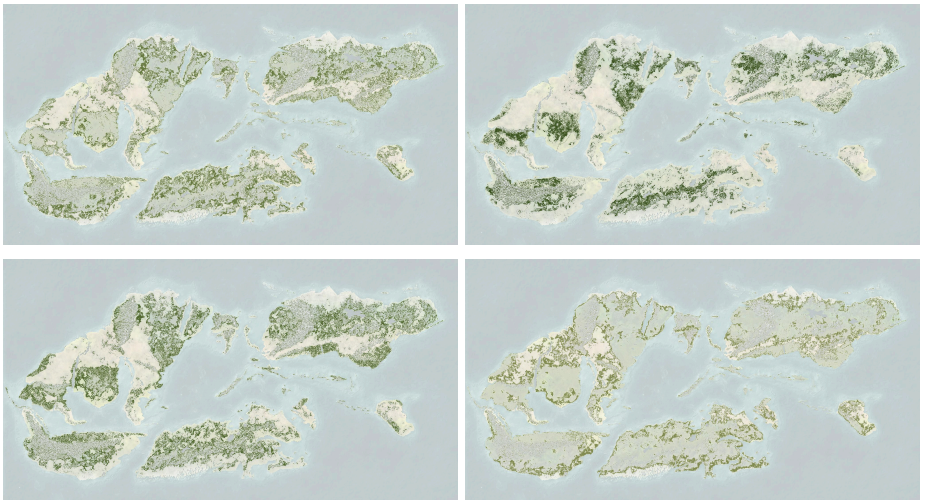
Les différences entre les trois normes semblent très minimes et se remarquent plus sur certains biomes comme l'océan, c'est peut-être dû au nombre de clusters trop bas qui empêchent de voir une certaine différence.

Essaie avec 20 cluster :

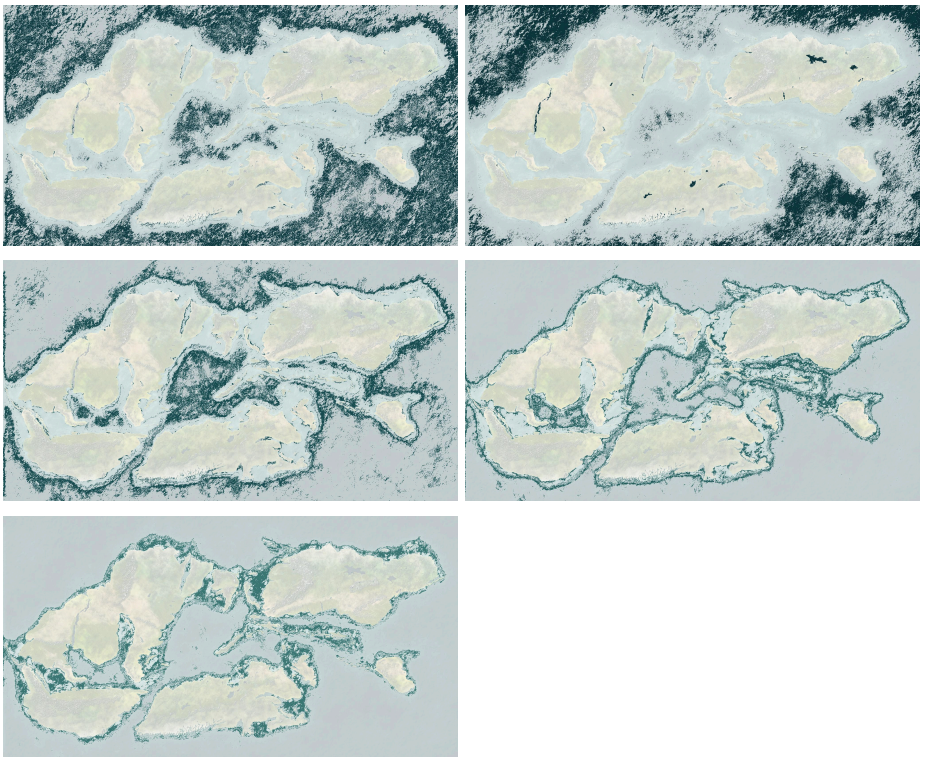
j'ai donc augmenté, pour l'image 3, le nombre de clusters testé pour chaque norme. Comme on peut s'y attendre les différences entre les couleurs sont beaucoup plus marquées et cela se remarque notamment pour les océans, là où à la première étape avec la norme normale, on ne voyait que 2 biomes pour les océans, on en voit bien plus maintenant, on voit aussi beaucoup plus de forêts différentes.

Normal :

-Forêts :

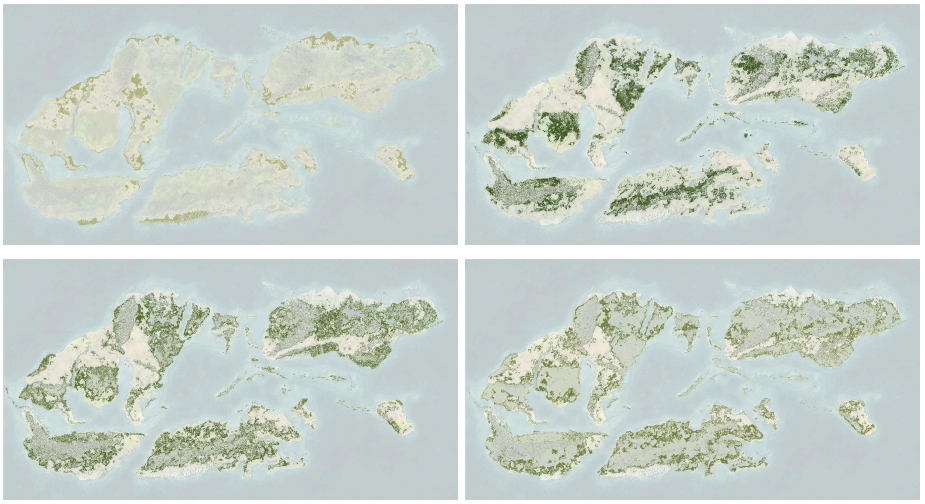


-Océan :

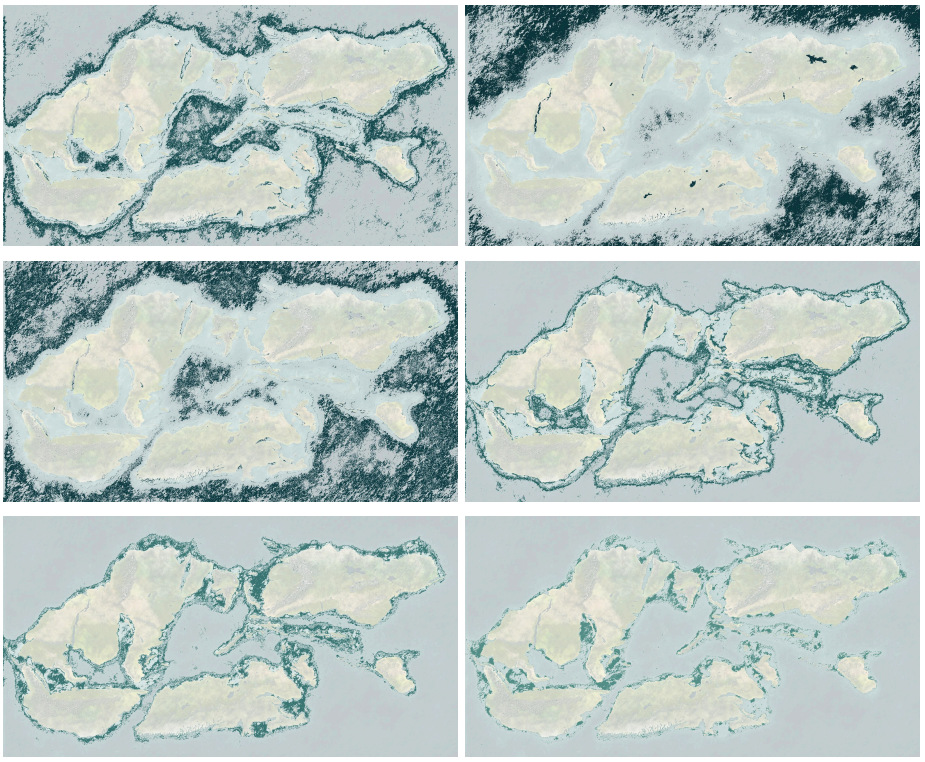


RedMean :

-Forêts :

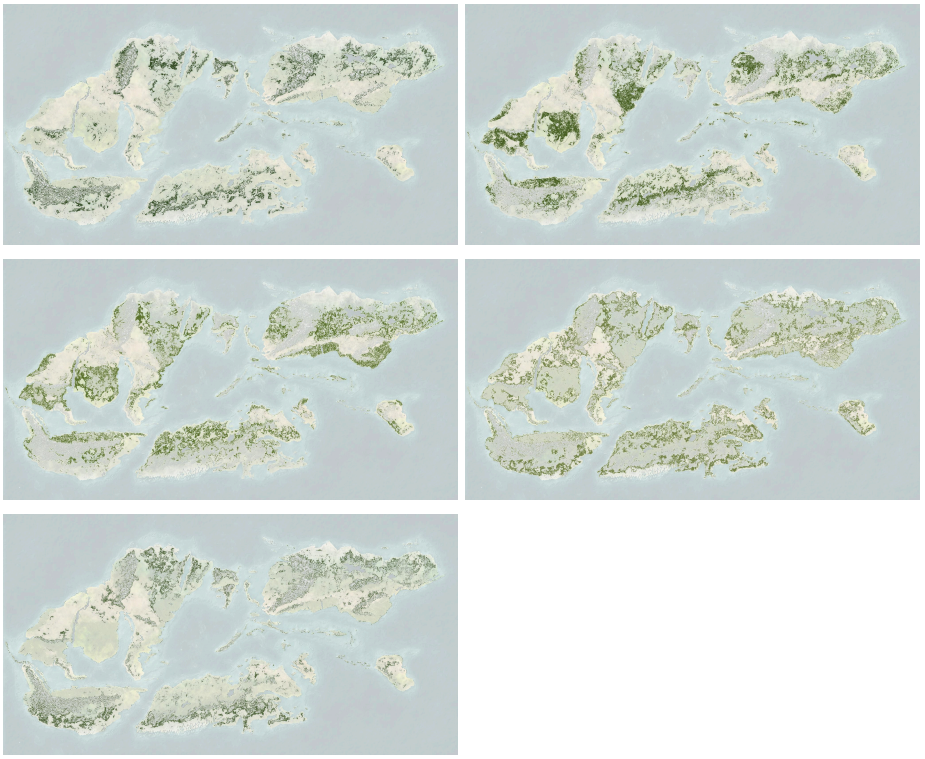


-Océan :

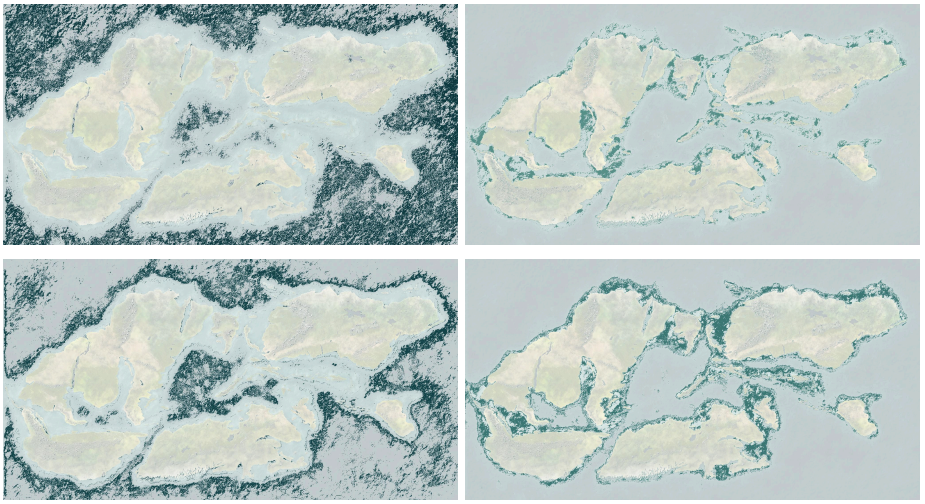


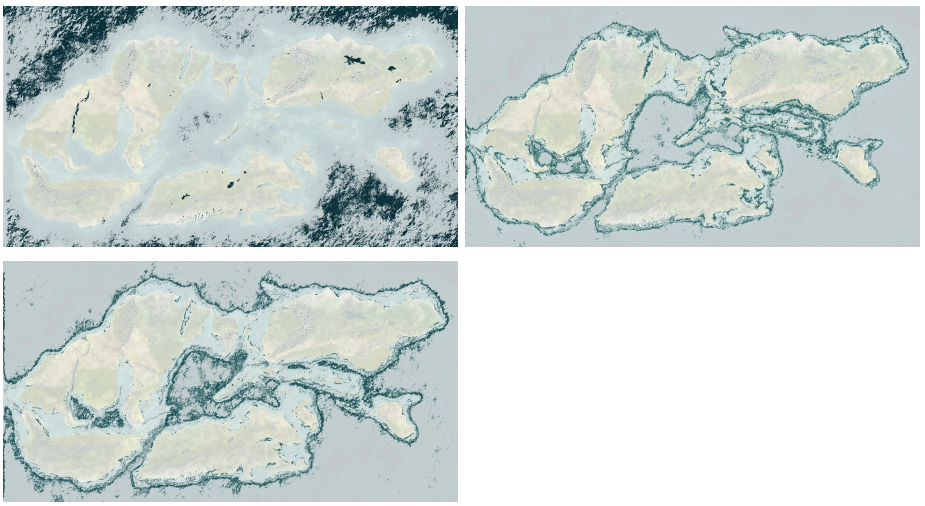
-CieLab :

-Forêts :



-Océan :

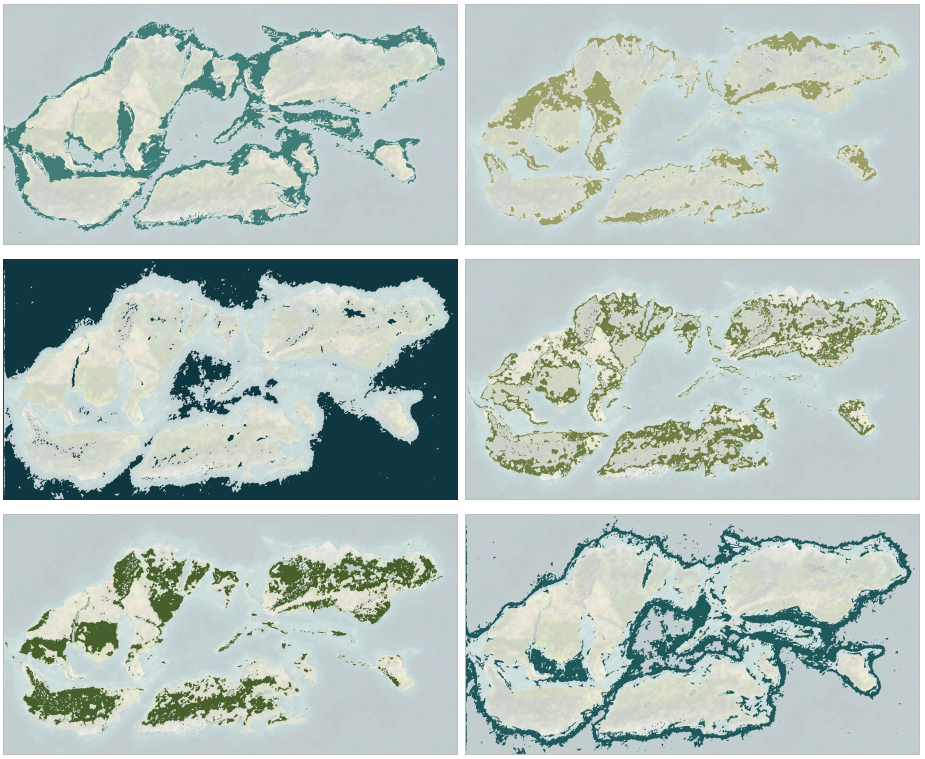




On peut aussi noter :
Bien que rajouter de nombreux cluster ralenti le programme et parfois nous donne des biomes inutiles, il permet de faire ressortir des biomes qu'on ne peut pas voir avec moins de cluster tel que le biome montagne :

Flou Gaussien :

On peut aussi, avant de détecter les biomes, appliquer un flou gaussien pour lisser l'image pour homogénéiser les couleurs.



On peut voir que l’image est beaucoup plus épurée et les biomes bien mieux unifier, on trouve moins de “trou” dans les biomes.

Meilleur norme :

D’après les résultats obtenus, nous en avons conclu que la norme à choisir était RedMean. En effet, c'est un bon compromis entre la norme Basique qui va plus vite pour trouver les différents biomes. RedMean est plus précis que la Norme Basique, prend en compte la perception humaine et va bien plus vite que la norme CieLab. Mais finalement toutes les normes ont leurs propres moyens de différencier les biomes. CieLab est définitivement la plus précise mais aussi la plus coûteuse.

3. Détection et affichage des écosystèmes pour chaque biome

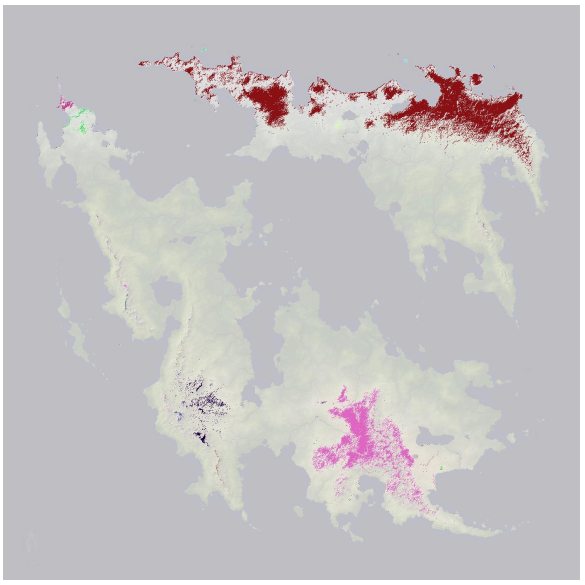
Nous avons choisi d'utiliser l'algorithme DBSCAN pour identifier les écosystèmes présents au sein d'un même biome, car il est difficile de connaître à l'avance le nombre exact d'écosystèmes. Il peut y avoir un seul écosystème ou bien une centaine, contrairement au nombre de biomes que l'on peut approximer à une dizaine par planète. DBSCAN est particulièrement adapté à cette tâche, parce qu'il n'exige pas de spécifier le nombre de clusters à l'avance et peut détecter des structures de données de forme irrégulière, ce qui est courant pour des écosystèmes.

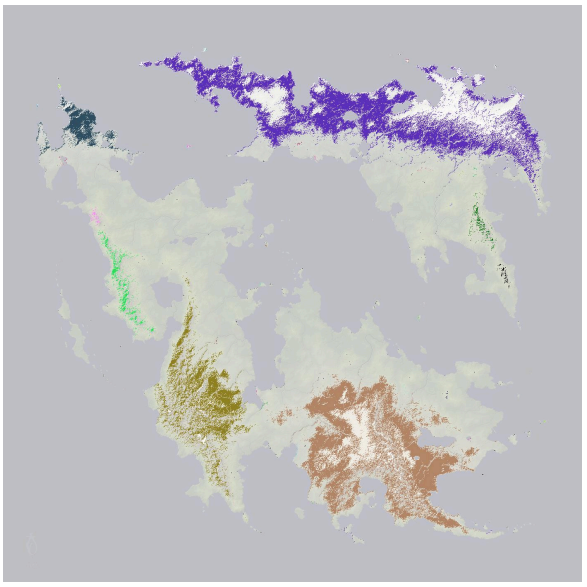
Pour représenter des sous-écosystèmes contenus dans d'autres, notamment dans des écosystèmes très denses avec plusieurs niveaux de densité, l'utilisation de l'algorithme HAC pourrait s'avérer particulièrement bénéfique.

Cependant, nous avons choisi de ne pas retenir HAC principalement en raison de sa complexité, qui peut être significative, surtout avec des images à haute résolution. Aussi, il est important de reconnaître que pour des images à très faible résolution ou pour des analyses où la hiérarchie des clusters est essentielle pour comprendre la structure des écosystèmes, HAC reste une bonne option.

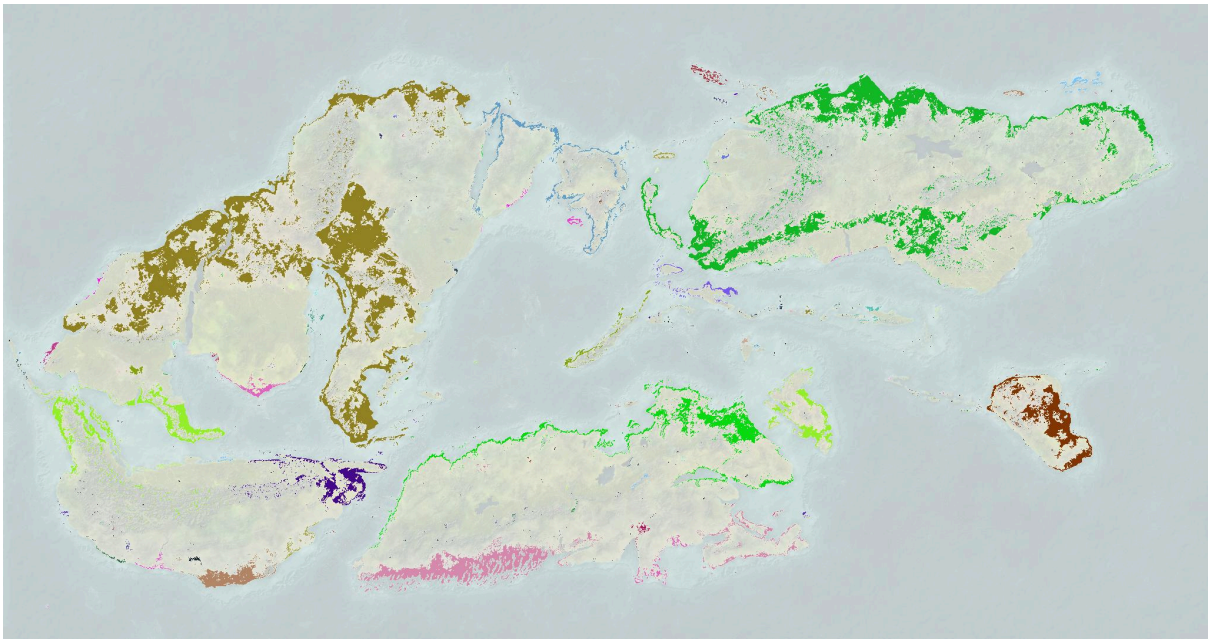
Images obtenus grâce au programme implémentant DBSCAN :

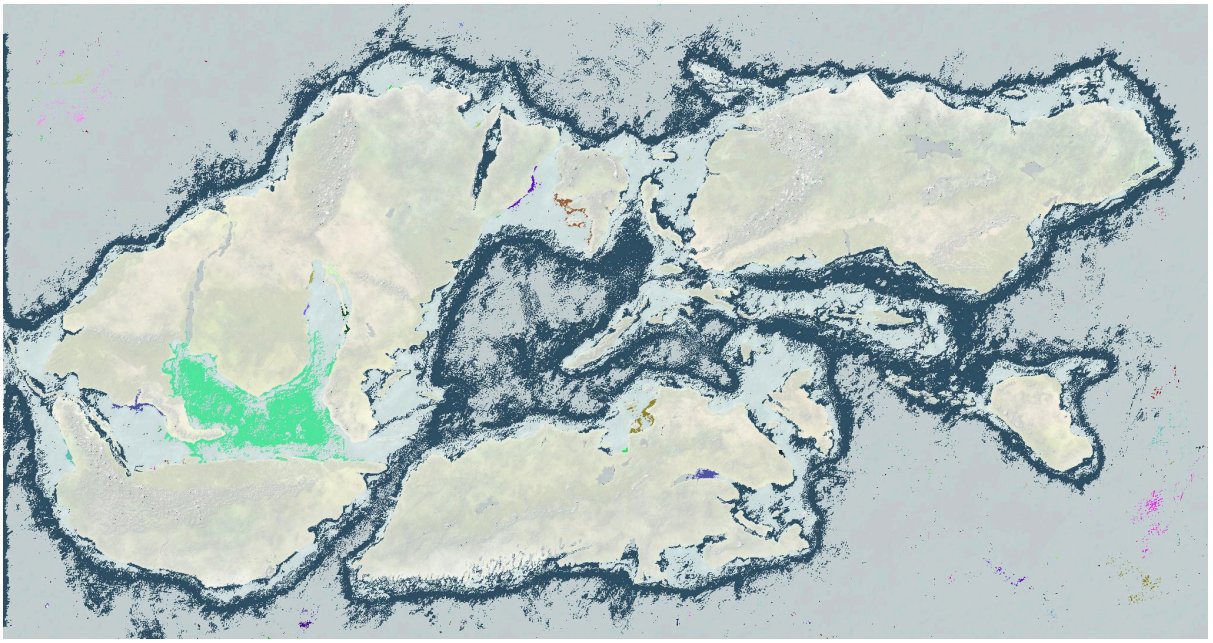
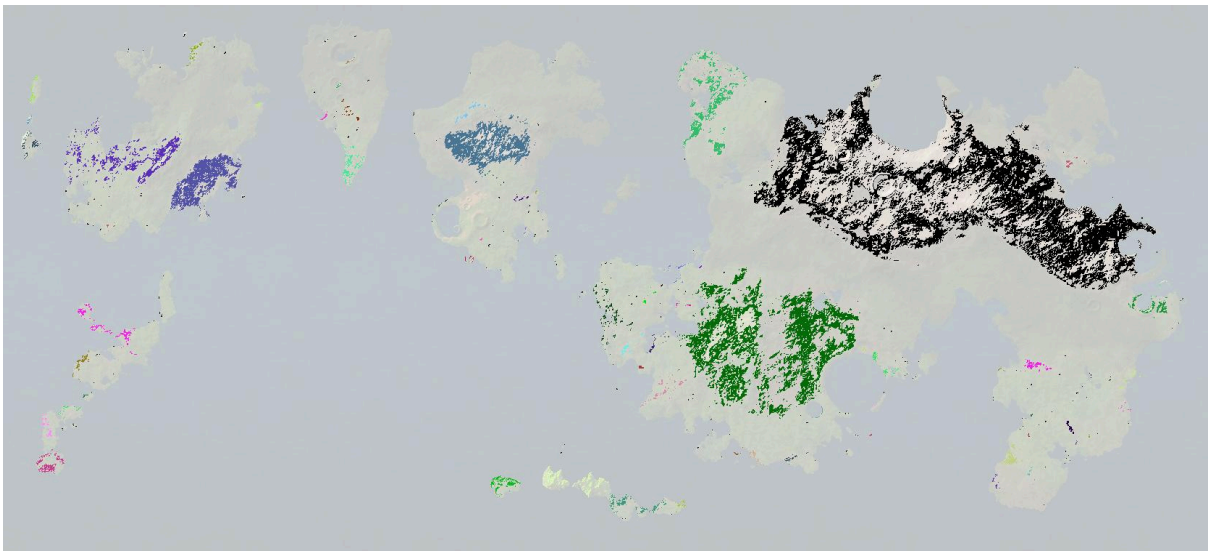
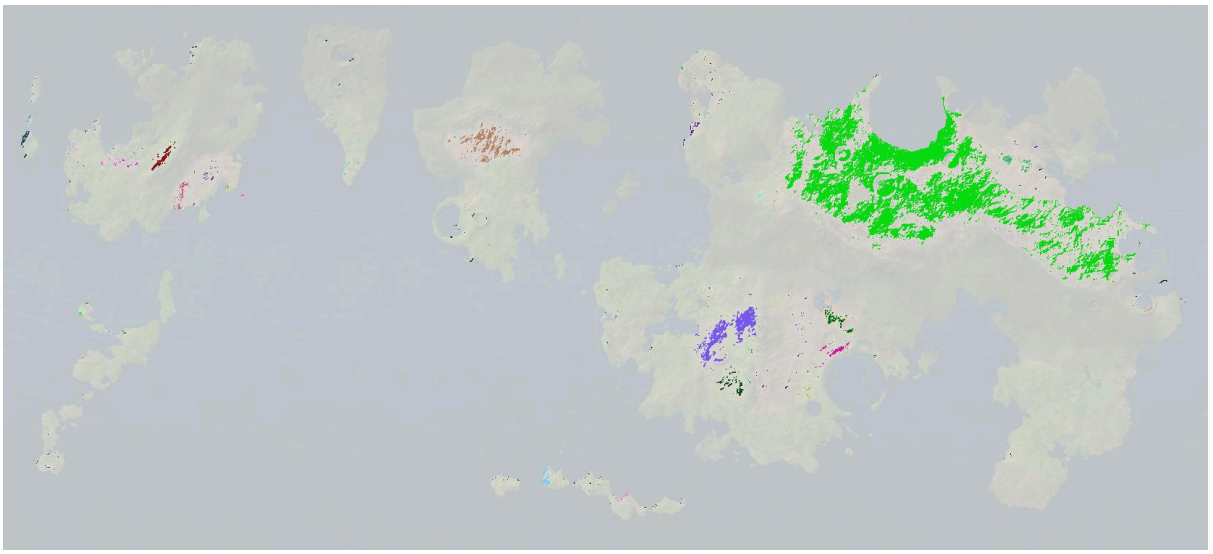
Planète 1 : RedMean





Planète 3 : RedMean





Planète 5 : Normal