

# Classification supervisée des images satellitaires pour Genève

---

Rapport de stage



Figure 1: <https://etggroupe.wordpress.com/2018/09/28/comprendre-ce-quest-la-resolution-spatiale/> (Pierro)

**Rédigé par**

Stauffer Anthony

**Superviseur**

M. Giuliani

**Aide**

M. Chognard



**UNIVERSITÉ  
DE GENÈVE**

## Remerciements :

---

Je tiens à remercier tout particulièrement M. Giuliani pour sa compréhension et sa flexibilité, qui m'ont permis de mener à bien ce stage tout en tenant compte de mes obligations académiques et professionnelles.

Je remercie également M. Sébastien Chognard pour son aide précieuse face aux problèmes techniques rencontrés durant ce stage.

Je remercie aussi M. Simon Carlier, qui m'a aidé sur certains aspects techniques lors de ce stage.

# Liste des abréviations :

---

- GRID : Global Resource Information Database
- NOLC04 : Nomenclature des occupations du sol Suisse 2004
- LCCS : Land Cover Classification System
- PNUE : Programme des Nations Unies pour l'Environnement
- CRS : système de coordonnées de référence
- WFS : Web Feature Service
- TERN : Terrestrial Environment Research Network
- FAO : l'Organisation des Nations Unies pour l'alimentation et l'agriculture
- SIG : système d'information géographique
- NDVI : Indice de végétation par différence normalisée
- RGB : bandes rouge, vert et bleu
- SVM : Support Vector Machine
- NDWI : Normalized Difference Water Index
- NDBI : Normalized Difference Built-up Index
- GCI : Green Chlorophyll Index

## Résumé :

---

L'objectif de ce projet était de réaliser une classification supervisée des images satellitaires de Genève en utilisant l'application EarthTrack, qui permet de géolocaliser et de classer des points à l'échelle mondiale. Ce travail s'inscrit dans une démarche d'amélioration des méthodes de cartographie du territoire suisse, en s'appuyant sur des images satellitaires à haute résolution.

La première étape a consisté à adapter EarthTrack à la nomenclature suisse NOLC04\_27, qui n'était pas initialement prise en charge par l'application, car celle-ci fonctionne par défaut avec le système de classification des couvertures terrestres de la FAO (LCCS). Une correspondance a été établie entre ces classifications afin de garantir la cohérence entre les données collectées sur le terrain et leur interprétation dans un système d'information géographique (SIG).

Une campagne de collecte de données a ensuite été menée à Genève, avec un total de 167 points géolocalisés et classifiés selon la nomenclature adaptée. Ces points ont été intégrés dans un processus de classification supervisée utilisant principalement le logiciel ArcGIS Pro, avec l'algorithme support vector machine (SVM). Deux sources de données satellitaires ont été comparées : Sentinel-2 (résolution de 10 m) et Landsat 8 (résolution de 30 m).

Les résultats montrent que les images Sentinel-2 permettent une classification plus précise et homogène que celles de Landsat 8, en particulier pour distinguer les surfaces urbanisées et les plans d'eau. Toutefois, plusieurs limitations ont été identifiées, notamment la nécessité d'un plus grand nombre de points de terrain pour améliorer la précision du modèle, ainsi que l'impact de la résolution des images sur la qualité de la classification.

Si les résultats démontrent la supériorité de Sentinel-2 pour une classification détaillée, certaines contraintes persistent, notamment en ce qui concerne l'adaptation d'EarthTrack à la nomenclature suisse et la disponibilité de données de terrain. Pour pallier ces limites, il serait pertinent d'explorer l'acquisition de données complémentaires, ainsi que l'intégration de nouvelles sources d'images satellitaires à plus haute résolution.

Ce travail met en lumière le potentiel de l'application EarthTrack pour la collecte participative de données et son intégration dans des processus de classification supervisée. Des améliorations sont envisagées, notamment l'extension du projet à un territoire plus vaste et l'exploration de nouvelles sources de données satellitaires à plus haute résolution.

# Table des matières :

---

<b>INTRODUCTION :</b>	6
<b>CONCEPTS THÉORIQUES MOBILISÉS :</b>	7
INTRODUCTION .....	7
PRÉSENTATION DES CONCEPTS .....	7
<i>Télédétection</i> .....	7
<i>Classification supervisée</i> .....	7
<i>Analyse spatiale</i> .....	7
<i>Systèmes d'information géographique (SIG)</i> .....	8
<i>Qualité des données</i> .....	8
<i>Application des concepts</i> .....	8
<b>DONNÉES ET MÉTHODOLOGIE</b> .....	9
DÉTAILS DES TÂCHES ET MISSIONS RÉALISÉES DURANT LE STAGE : .....	9
ÉTUDE ET TEST DE L'APPLICATION EARTHTRACK .....	9
ADAPTATION CLASSIFICATION SUISSE NOLC04_27 .....	10
TRADUCTION ET MISE EN CORRESPONDANCE DES DÉFINITIONS EARTHTRACK AVEC NOLC04_27 .....	10
COLLECTE DE DONNÉES À GENÈVE .....	10
MACHINE LEARNING .....	11
<i>Système d'information géographique (SIG)</i> .....	11
<i>Téléchargements donnés</i> .....	12
<i>Composition des bandes</i> .....	12
<i>Délimitation Canton de Genève</i> .....	13
<i>Importation de données dans ArcGIS Pro</i> .....	13
<i>Classification supervisée</i> .....	13
<i>Layout</i> .....	16
<b>PRÉSENTATION ET ANALYSE DES RESULTATS</b> .....	20
APPORT POTENTIEL DES DONNÉES LiDAR POUR AFFINER LA CLASSIFICATION .....	21
<b>DISCUSSION ET CONCLUSIONS</b> .....	24
LIMITES DE L'ÉTUDE .....	24
EFFICACITÉ D'EARTHTRACK .....	25
RECOMMANDATIONS ET PERSPECTIVES D'AVENIR .....	26
REFLEXIONS SUR LE DÉROULEMENT DU STAGE .....	27
<b>ANNEXE:</b> .....	28
DÉFINITION EARTHTRACK: .....	28
TABLEAU EARTHTRACK x NOLC04_27 .....	31
SCRIPT : .....	47
<b>BIBLIOGRAPHIE :</b> .....	48

## Introduction :

Dans un contexte où l'actualisation des données cartographiques est essentielle pour une gestion optimale des ressources et du territoire, l'exploitation des images satellitaires s'avère prometteuse. Cependant, les cartes suisses utilisées pour la classification du territoire ont une résolution spatiale de 100 m, ce qui limite la précision des analyses et des décisions basées sur ces données. Cette faible résolution peut avoir des répercussions sur plusieurs domaines, notamment la planification urbaine, la gestion des milieux naturels et le suivi environnemental. Afin de pallier cette limitation, l'objectif de ce projet est d'évaluer l'utilisation de l'application EarthTrack pour améliorer la classification du territoire à l'aide de techniques de télédétection et de systèmes d'information géographique (SIG).

Comme l'ont souligné Chaturvedi et de Vries (2021), les méthodes traditionnelles de cartographie et d'analyse du territoire ont du mal à suivre face à la complexité des dynamiques urbaines. Les approches classiques sont souvent statiques et peinent à saisir les changements imprévisibles et complexes de l'utilisation des sols. Dans ce contexte, les techniques d'apprentissage automatique (machine learning), notamment les algorithmes de classification supervisée (SVM), offrent une alternative prometteuse pour améliorer la précision des classifications et la compréhension des dynamiques territoriales (Chaturvedi & de Vries, 2021).

L'objectif principal de ce travail est d'évaluer dans quelle mesure l'application EarthTrack, couplée à une classification supervisée sur des images Sentinel-2 et Landsat 8, permet d'améliorer la cartographie du territoire genevois. Nous formulons l'hypothèse que Sentinel-2, grâce à sa meilleure résolution spatiale et spectrale, permet une classification plus précise que Landsat 8.

L'utilisation d'EarthTrack présente plusieurs avantages : elle permet une collecte collaborative des données par les utilisateurs, avec validation par des experts. Toutefois, son utilisation en Suisse se heurte à un défi majeur : l'application utilise par défaut le système de classification des couvertures terrestres de la FAO (LCCS), qui ne correspond pas directement aux classifications employées dans ce pays. Ce projet vise donc à adapter EarthTrack à la nomenclature suisse NOLC04\_27, à tester la classification supervisée des images satellitaires, et à évaluer la pertinence de cette approche pour la mise à jour cartographique du territoire.

L'expérimentation a été menée sur le canton de Genève, qui a joué le rôle de territoire pilote. Un ensemble de données a été collecté sur le terrain, puis intégré à des logiciels SIG afin d'entraîner un modèle de classification supervisé. Deux types d'images satellitaires ont été comparés : Sentinel-2 (résolution de 10 m) et Landsat 8 (résolution de 30 m). L'objectif est d'analyser la qualité des classifications obtenues, et de déterminer dans quelle mesure EarthTrack peut contribuer à améliorer les données géospatiales en Suisse.

Ce rapport commence par présenter le contexte et les enjeux du projet, puis détaille les concepts théoriques mobilisés et la méthodologie employée. Les résultats obtenus feront ensuite l'objet d'une analyse et d'une discussion, afin d'identifier les forces et les limites de cette approche, et de proposer des pistes d'amélioration pour de futures recherches.

# Concepts théoriques mobilisés :

## Introduction

Dans le cadre de cette étude, plusieurs concepts théoriques issus de la télédétection, de l'analyse spatiale et du machine learning ont été mobilisés pour structurer et guider la recherche. Ces concepts fournissent un cadre de référence essentiel pour formuler les hypothèses, analyser les données et interpréter les résultats de manière cohérente.

## Présentation des concepts

### Télédétection

La télédétection est la science de l'acquisition et de l'interprétation des données obtenues à distance, généralement par des satellites ou des avions (Université de Genève, 2024). Dans notre cas, nous avons utilisé des images satellitaires Landsat 8 ainsi que Sentinel-2. La télédétection permet de comprendre comment les capteurs recueillent des informations sur la surface terrestre et comment ces données peuvent être utilisées, par exemple, pour l'analyse environnementale ou territoriale.

### Classification supervisée

La classification supervisée est une technique de machine learning utilisée pour catégoriser les données en fonction de classes prédéfinies. Comme le soulignent Chaturvedi et de Vries (2021), les algorithmes de classification supervisée, tels que les support vector machines (SVM), sont particulièrement efficaces pour l'analyse des images satellitaires. Ces algorithmes permettent de modéliser des relations complexes entre les variables spectrales et les classes de couverture du sol, tout en minimisant les erreurs de classification. Dans le cadre de mon stage, nous avons utilisé cette méthode pour classer les différentes zones de couverture terrestre à Genève. La classification supervisée implique l'utilisation d'un algorithme entraîné sur des données d'apprentissage pour attribuer des étiquettes de classe à de nouvelles données. L'algorithme fonctionne de la manière suivante : parmi un lot de 10 points à classer, il en prend 5 ou 6 pour établir les règles de classification, et il en prend 4 ou 5 pour la vérification des autres points.

### Analyse spatiale

L'analyse spatiale consiste à examiner les relations spatiales entre les objets à l'aide de techniques géospatiales. Les concepts liés aux systèmes de coordonnées de référence (CRS) et aux techniques de géotraitement sont essentiels pour garantir la précision et la cohérence des analyses géospatiales. Dans le cadre de mon stage, nous avons utilisé des techniques d'analyse spatiale pour étudier la distribution des différentes classes de couverture terrestre à Genève. Nous avons principalement eu recours aux techniques suivantes : la superposition de couches, la création de buffers et l'analyse de voisinage. Il est très important d'utiliser des CRS appropriés pour garantir la précision et la cohérence des analyses géospatiales, car sinon l'alignement et la superposition des couches peuvent être affectés, entraînant une perte de fiabilité des résultats.

## Systèmes d'information géographique (SIG)

Les SIG sont des systèmes informatiques puissants conçus pour la gestion et l'analyse des données géospatiales. Leurs fonctionnalités clés incluent la gestion des données spatiales, la visualisation cartographique, l'analyse spatiale et la modélisation. Les principaux SIG utilisés dans ce rapport sont ArcGIS Pro et QGIS. ArcGIS Pro est un logiciel payant, tandis que QGIS est gratuit.

## Qualité des données

La qualité des données géospatiales dépend de leur précision, de leur résolution et de leur fiabilité. Dans le cadre de ce stage, nous avons constaté que la résolution des images satellitaires avait un impact significatif sur la qualité des résultats. La qualité des données est un aspect essentiel de toute analyse géospatiale, et l'objectif de ce stage était justement d'observer l'impact d'une meilleure résolution des cartes. Les données de mauvaise qualité peuvent en effet conduire à de faux résultats ainsi qu'à des décisions inappropriées. Plus la résolution est élevée, plus les images permettent d'identifier des objets de plus en plus petits et d'obtenir des classifications de plus en plus précises.

## Application des concepts

Dans cette étude, les concepts théoriques de la télédétection, de la classification supervisée, de l'analyse spatiale, des SIG et de la qualité des données ont été appliqués pour développer une méthodologie de classification supervisée des images satellitaires de Genève. Les théories de la télédétection ont, par exemple, guidé le choix des images satellitaires, tandis que les modèles de classification supervisée ont servi à développer les algorithmes de classification.

## Données et méthodologie

### Détails des tâches et missions réalisées durant le stage :

- ⇒ Il a tout d'abord fallu lire l'article *Living Earth, EarthTrack and TERN*, fourni par M. Giuliani, qui présente le fonctionnement de l'application EarthTrack ainsi que des cas d'usage concrets.
- ⇒ Ensuite, on a dû se familiariser avec l'application.
- ⇒ On a dû créer un cheminement pour la classification suisse à l'aide de cette application.
- ⇒ Des photos géolocalisées ont ensuite été prises sur le terrain à Genève, avec environ dix observations par classe, à l'aide de l'application EarthTrack.
- ⇒ Ces points ont ensuite été importés dans un environnement SIG, puis exploités à l'aide de techniques de machine learning pour produire des cartes de classification du territoire.
- ⇒ Enfin, une évaluation a été menée pour déterminer si EarthTrack représente un outil pertinent pour la classification territoriale à l'échelle d'un canton comme Genève, et, le cas échéant, s'il serait envisageable d'en étendre l'usage à une échelle plus large.

### Étude et test de l'application EarthTrack

Pour commencer, nous nous sommes focalisés sur l'article descriptif que M. Giuliani m'a fourni concernant l'application EarthTrack et son partenariat avec le TERN australien. EarthTrack est une application liée à « Living Earth », un système qui crée des cartes de couverture terrestre à partir de données observées. L'application EarthTrack permet à tout un chacun d'enregistrer des informations sur la couverture terrestre et les changements environnementaux sur le terrain, ce qui permet d'affiner les cartes thématiques de Living Earth. Ce rapport montre surtout que EarthTrack est un bon outil pour la collecte de données standardisées sur la couverture terrestre et les changements environnementaux. Il est donc parfaitement adapté à notre projet.

Ce constat est également soutenu par le rapport de TERN (2023), qui met en avant l'intérêt de l'application EarthTrack pour standardiser la collecte de données environnementales sur le terrain, tout en soulignant certaines limites techniques, telles que la stabilité des serveurs ou la dépendance à une bonne connectivité.

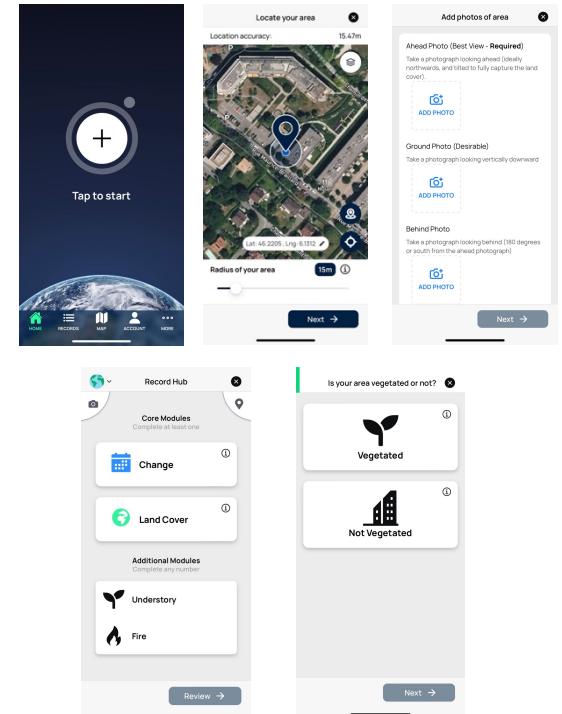


Figure 2 : Screenshot de l'application EarthTrack qui montre son interface ainsi que son l'interactivité

À noter que Living Earth utilise le système de classification de la couverture terrestre de la FAO (LCCS). Pour simplifier, EarthTrack utilise également le même LCCS. Pour notre part, nous voulons utiliser le NOLC04\_27, ce qui nécessitera une adaptation.

Concernant la première prise en main, celle-ci est très intuitive et interactive, ce qui facilite grandement la manipulation. En effet, l'application n'est pas surchargée d'informations et le fait qu'elle soit interactive aide beaucoup à cela. À chaque étape, il est possible de progresser en fonction de différentes catégories, avec une interface bien guidé (voir figure 2).

## Adaptation classification Suisse NOLC04\_27

Concernant l'adaptation de la classification suisse NOLC04\_27, nous avons commencé par lire la documentation correspondante, puis tenté de déterminer le cheminement adéquat pour chaque critère associé à chaque classe (voir tableau en annexe). De plus, pour m'aider dans cette tâche, nous avons traduit tous les termes proposés par l'application EarthTrack pour chaque cheminement (voir définition EarthTrack en annexe) (Office fédéral de la statistique, 2019).

Il est à noter que cette adaptation ne suit pas toujours un modèle de correspondance de type un pour un.

En effet, plusieurs classes de la nomenclature suisse peuvent partager un même cheminement FAO du LCCS dans EarthTrack. Cela signifie qu'il n'existe pas nécessairement un chemin unique entre chaque classe suisse et les catégories proposées dans EarthTrack. Certaines classes ont été regroupées parce qu'elles sont proches ou similaires.

Cette démarche a permis de rationaliser l'utilisation de l'application tout en maintenant une compatibilité acceptable avec la nomenclature suisse, bien que certaines spécificités locales aient dû être simplifiées.

## Traduction et mise en correspondance des définitions EarthTrack avec NOLC04\_27

Pour ce faire, nous avons suivi tous les cheminements proposés et, pour parvenir à la dernière sélection, ce sont ces termes-là que nous avons traduits. Un petit « i » apparaît au-dessus de chaque terme, ce qui aide à comprendre plus précisément leur signification. Cette traduction permet donc de mieux comparer les critères de la documentation de la classification pour chaque catégorie et le cheminement d'EarthTrack.

Ce cheminement nous permet finalement de savoir exactement quel parcours suivre pour catégoriser une classe souhaitée, ce qui représente un gain de temps considérable.

## Collecte de données à Genève

Grâce au cheminement préalablement établi, nous sommes allés un peu partout dans Genève pour pouvoir récolter des données. En effet, pour que les concepteurs de l'application puissent valider les points, il faut fournir une photo. C'est Sébastien Chognard qui s'en charge actuellement. La collecte de données consistait donc à visualiser au préalable les images satellitaires disponibles, comme celles de Google

Cancel	Life form
Shrubs	(i)
Graminoids	(i)
Non-graminoids	(i)

**Non-graminoids** 

All herbaceous plants that are not graminoids (i.e., species that do not belong to the grasses according to the taxonomic definition). Can cover large surface areas of water if aquatic.

Figure 3 : Exemple d'explication d'un terme sur l'application EarthTrack, ici les non-graminoides

Maps, pour localiser une zone et déterminer quels points pouvaient être pris en compte pour la classification dans cette zone avant de s'y rendre. Finalement, nous avons constaté qu'il existe 16 catégories parmi les 27 qui sont exploitables à Genève. (Office fédéral de la statistique, 2019). Les 16 classes retenues pour Genève sont :

1. Consolidated surface | (Surfaces compactées)
2. Buildings | (Bâtiments)
3. Greenhouses | (Serres)
4. Gardens | (Structures de cultures en planches)
5. Lawns | (Gazon)
6. Trees in artificial areas | (Arbres sur terrains aménagés)
7. Mix of small structures | (Petites structures mixtes)
8. Grass & herbe vegetation | (Végétation herbacée)
9. Short stem fruit trees | (Arbres fruitiers à basses tiges)
10. Vines | (Vignes)
11. Permanent garden plants & brush crops | (Cultures horticoles permanentes)
12. Closed forest | (Peuplements d'arbres denses)
13. Forest Strip | (Bandes de forêt)
14. Linear woods | (Peuplements d'arbres linéaires)
15. Water | (Plans d'eau)
16. Reedy marshes | (Roselières)

Nous avons décidé de ne conserver que ces catégories, car pour les autres, soit il n'y avait pas suffisamment d'espace (à notre connaissance) pour pouvoir recueillir assez de données, soit il était trop compliqué de distinguer une autre catégorie à une si petite échelle qu'est le canton de Genève, ou bien la topographie du canton ne possède pas cette classe.

Une fois sur place, nous choisissons le « rayon » de la zone, nous prenions une photo pour valider le point et on suivait le cheminement préalablement défini pour chaque classe. Enfin, nous inscrivions toujours « UNIGE, numéro de la classe\_Nom de la classe » dans les commentaires, afin de pouvoir se retrouver plus facilement par la suite lors du processus de machine learning avec un système d'information géographique.

Pour la collecte de données, nous utilisions un vélo pour nous déplacer d'un endroit à l'autre, en procédant par région (inconvénient : l'autonomie de ma batterie était faible ; on avait le temps de faire 15 à 20 photos par sortie).

## Machine Learning

### Système d'information géographique (SIG)

Après la collecte des données, il faut les traiter. Nous avions tout d'abord convenu avec M. Giuliani de travailler avec QGIS pour plusieurs raisons. La première est que c'est un logiciel gratuit et téléchargeable sur tous les appareils (iOS, Windows et Linux), contrairement à d'autres logiciels comme ArcGIS Pro, qui ne sont pas disponibles sur iOS.

Il existe également un plugin, le « Semi-Automatic Classification Plugin », sur QGIS, qui permet de réaliser des classifications à l'aide de techniques de machine learning.

Malheureusement, nous avons rencontré plusieurs problèmes avec ce logiciel. En effet, nous travaillons sous macOS, avec un ordinateur relativement ancien, ne permettant plus d'effectuer de mises à jour. Pour installer le plugin, il faut suivre un protocole spécifique, et la version de QGIS compatible avec cet ordinateur ne supportait pas le plugin. Un message d'erreur apparaissait à chaque tentative. Nous avons donc abandonné l'idée de travailler sur l'ordinateur personnel, et avons poursuivi sur les ordinateurs de l'université, mais à distance. Cette fois-ci, tout a bien fonctionné, mais le traitement était bien trop lent. Nous avons donc finalement choisi d'utiliser directement les postes fixes à l'université. L'inconvénient majeur avec ces ordinateurs est qu'ils se réinitialisent chaque jour : il faut donc réinstaller le plugin à chaque redémarrage, même si les données sont bien stockées sur une clé USB (le logiciel n'étant pas installé sur cette dernière). (Congedo, 2024).

## Téléchargements donnés

Une fois le plugin installé, il fallait télécharger les données stockées sur Living Earth. Nous avons contacté Sébastien Chognard pour qu'il nous indique comment se les procurer et il nous a dit qu'on pouvait les obtenir directement via le lien « <https://earthtrack.aber.ac.uk/download/> ». Malheureusement, à chaque fois, des erreurs survenaient lorsqu'on essayait de les télécharger et Sébastien nous a indiqué que le serveur était en panne. Il nous a donc envoyé lui-même les données au format GeoJSON, après avoir filtré les notres (grâce au champ « commentaires » qui commençait par « UNIGE » (Aberystwyth University, n.d.)).

## Composition des bandes

Nous avons décidé de travailler avec Landsat 8 pour commencer, sachant qu'il offre une résolution de 30 mètres, ce qui est bien meilleur que les 100 mètres actuellement utilisés en Suisse. On a donc téléchargé les bandes sur EarthExplorer, le 29 août 2024, période durant laquelle les photos ont été prises – où l'on distingue bien la végétation – ce qui nous a été utile par la suite pour la composite fausse couleur NIR (USGS, n.d.). Nous avons également sélectionné une date où il n'y avait pratiquement pas de nuages à Genève pour éviter toute interférence. Le plugin ne fonctionnait pas correctement pour réaliser des compositions de bandes. Après plusieurs heures d'essais et de lecture de forums pour résoudre le problème, on a décidé de passer sur ArcGIS Pro, car on trouvait qu'il y avait trop d'inconvénients à travailler sur QGIS.

Nous avons donc procédé à la composition des bandes sur ArcGIS Pro. Après avoir inséré mes bandes dans ArcGIS Pro, on a utilisé l'outil « Composite Bands » en sélectionnant les bonnes bandes pour obtenir une composition en fausses couleurs RGB-NIR, respectivement en vert et en bleu. Nous avions oublié de vérifier une erreur : le système de coordonnées de référence (CRS) doit être identique pour chaque couche. Nous avons donc corrigé cela à l'aide des images Landsat 8 téléchargées sur EarthExplorer, en utilisant le CRS CH1903+ LV95 pour toutes les couches (USGS, n.d.). Depuis lors, on a toujours fait attention à cet aspect pour toutes les couches.

## Délimitation Canton de Genève

Nous avons ensuite récupéré le contour de Genève sur le site des SITG, afin de créer un clip des bandes en fonction des limites du canton. Dans ArcGIS Pro, nous avons utilisé l'outil « Feature to Polygon » pour créer une couche de délimitation du canton de Genève, que nous avons ensuite utilisée pour découper les bandes (SITG, n.d.). Lors de l'utilisation de l'outil « Clip Raster », il ne faut pas oublier de cocher la case « Use Input Features for Clipping Geometry ». Une fois les bandes Landsat 8 découpées selon la forme du canton de Genève, nous avons réalisé une composition à partir de ces bandes.

## Importation de données dans ArcGIS Pro

Nous avons donc voulu importer les points dans ArcGIS à partir du fichier GeoJSON que Sébastien nous avait envoyé contenant l'ensemble des points, mais ArcGIS ne reconnaissait pas ce format. On a donc ouvert ce fichier dans QGIS et nous l'avons exporté au format GeoPackage (.GPKG). Nous avons alors pu intégrer les points dans ArcGIS. Sébastien nous a ensuite conseillé de créer un nouveau champ pour simplifier le champ « comments » que l'on utiliserait pour la classification, et ce que nous avons fait.

## Classification supervisée

Au début de la classification, nous pensions qu'il était nécessaire d'utiliser les photos que nous avions prises pour valider les points géolocalisés sur EarthTrack. Nous avons donc passé beaucoup de temps à essayer d'associer chaque image à son point correspondant dans ArcGIS, mais sans succès. Après avoir consulté Sébastien, nous avons réalisé que cela n'était finalement pas nécessaire.

Le champ principal utilisé est « UNIGE\_Numéro de la classification\_Nom de la classification ». À ce moment-là, nous disposions de 80 points au total. Une partie a été utilisée pour entraîner l'algorithme de classification, tandis qu'une autre a été conservée pour valider la classification. Nous avons alors réalisé que nous ne disposions pas assez de photos par catégorie, car nous n'avions pas anticipé les points de validation lors de la collecte des données. On a donc décidé de continuer avec les points disponibles pour comprendre le fonctionnement du machine learning, si jamais on rencontrait des problèmes, et d'être en mesure de les résoudre. Une fois les données supplémentaires obtenues, il suffira de suivre exactement le même processus pour assurer le bon déroulement de l'opération.

Pour évaluer les performances du modèle, une matrice de confusion a été générée à partir des mêmes points utilisés pour l'entraînement.

Cette approche n'est pas idéale d'un point vue méthodologique, car elle combine les données d'entraînement et de validation, ce qui peut biaiser les résultats.

En temps normal, on sépare les jeux de données (train/test) pour afin d'éviter tout risque d'auto-validation.

Cependant, le faible nombre de points disponibles, les contraintes de temps et surtout les limitations techniques n'ont pas permis de constituer un échantillon distinct pour la validation.

Ensuite, nous avons voulu effectuer une classification supervisée via « Training Samples Manager », mais les véritables problèmes sont alors apparus. Il m'était impossible d'effectuer une classification supervisée avec les points dont nous disposions. Nous avons essayé pendant plusieurs jours, puis avons de nouveau demandé de l'aide à Sébastien. Nous avons convenu qu'il fallait ajouter, en plus des autres champs, un champ « classname » ainsi qu'un champ « classvalue » dans la table attributaire de la couche de points, pour que tout fonctionne. Mais cela ne suffisait pas : il nous fallait des polygones, et ArcGIS ne reconnaissait pas notre couche de points comme telle.

Nous devions également créer de nouvelles classes pour utiliser l'outil « **Image Classification** » via « **Training Samples Manager** ». Nous avons alors créé un fichier nommé « ClassSwiss.ecs » dans lequel on a enregistré les 16 classes différentes que nous avions retenues parmi la NOLC04\_27 pour Genève, dans l'outil « **Image Classification** », puis nous l'avons exporté afin de pouvoir l'utiliser ultérieurement. ClassSwiss.ecs contient donc les classes nécessaires à la classification supervisée. Malheureusement, cela ne nous a pas été d'une grande utilité, car il était impossible d'associer nos points aux classes de ClassSwiss.ecs.

Après cela, nous avons décidé de créer un « buffer » pour les points, en fonction du champ « radius ». Comme ce champ n'apparaissait pas dans la table attributaire, Sébastien nous a dit de créer un script Python pour l'intégrer. Nous avons donc utilisé un script Python (voir annexe) pour extraire cette information et la placer dans la propriété « radius » afin qu'elle apparaisse dans la table attributaire. Après cela, nous avons utilisé l'outil « Create Buffer » avec le champ « radius » pour définir la taille du buffer et on a désactivé l'option « Dissolve ».

Malgré tous ces efforts, nous nous sommes retrouvés bloqués au même stade... Après un nouveau rendez-vous avec Sébastien Chognard, nous avons réalisé que nous avions mal effectué la manipulation depuis le début. Il suffisait en fait de cliquer sur « **Classify** » (dans « **Classification Tools** » sous « **Imagery** »), car EarthTrack avait déjà effectué tout le travail en amont, et ça fonctionne ! il n'était donc pas nécessaire de passer par le « **Training Samples Manager** », qui ne fonctionne d'ailleurs pas pour ce cas. Mais tout le travail effectué auparavant n'était pas inutile, car le processus permettant d'obtenir le champ radius dans la table attributaire et de créer un buffer était indispensable pour pouvoir utiliser l'outil « **Classify** ».

Une fois que nous avons compris comment faire, nous sommes allés récolter des données pour compléter celles qu'on avait déjà.

Nous avons également procédé de la même manière avec Sentinel-2, afin d'obtenir une couche avec une meilleure résolution (10 m), et de pouvoir ainsi comparer les deux. Les images ont été téléchargées directement sur Copernicus (*Copernicus, n.d.*). On a veillé à sélectionner les bonnes bandes RGB pour obtenir la même fausse couleur qu'avec les images Landsat 8. Nous avons également mis ces bandes sous le bon format de CRS. Nous avons introduit les nouveaux points au format gpkg et on a également changé le CRS via l'outil « **Project** » en CH1903+\_LV95 (comme pour toutes les autres couches). Les données sont composées de 167 points au total. En moyenne, nous avons 10 points par catégorie, car on en a pris 16 au total.

Une fois le buffer appliqué à l'ensemble des points, lors du changement de symbologie, on a remarqué que certaines classes apparaissaient plusieurs fois (primary symbologie = Unique Values et field 1 = lc\_comments). On a donc vu que certains commentaires

n'étaient pas identiques, c'est-à-dire qu'il y avait par exemple une majuscule dans « **B**uilding » et une autre classe « Building » qui n'avait pas la majuscule « **b**uilding »). On a corrigé pour que tout soit uniforme et on a une seule classe par catégorie. Il fallait donc veiller à ce que toutes les entrées que l'on souhaite regrouper soient écrites de manière strictement identique dans la table attributaire.

Nous avons donc sélectionné la couche composite que l'on a créée à partir de Sentinel-2, puis on a utilisé l'outil de « classification tools » et on a mis « classify ». Dans « classify », on a choisi « Support Vector Machine » pour le « classifier » et, pour les échantillons d'entraînement, on a sélectionné la couche avec les buffers, qui se nomme « unige CRS Buffers » et qui comprend toutes les données/points. Nous avons laissé les autres paramètres par défaut.

Nous avons choisi l'algorithme Support Vector Machine pour la classification supervisée en raison de ses performances démontrées dans la littérature. Selon Chaturvedi et De Vries, les SVM sont particulièrement adaptés à la classification des images satellitaires, car ils permettent de gérer des données complexes et de modéliser des relations variées entre les variables spectrales. De plus, les SVM sont robustes face aux problèmes de surapprentissage, ce qui en fait un choix privilégié pour les études de télédétection où les données d'entraînement peuvent être limitées (*Chaturvedi & de Vries, 2021*).

On a alors reçu le message d'erreur suivant : "Error 002088 : Feature class has not enough information for supervised training."

On a dû ajouter un champ « class\_value » dans la table attributaire pour que le programme fonctionne à nouveau. Le problème persistait de façon aléatoire, nécessitant parfois un redémarrage du programme. Finalement, nous avons réussi à obtenir une classification supervisée. On a procédé exactement de la même manière avec les images Landsat 8.

Pour évaluer la précision de la classification supervisée, on a utilisé l'outil « Create Accuracy Assessment Points ». En entrée, on a sélectionné la couche classifiée Sentinel-2, et on a nommé l'output « Classified\_Sen\_Randompoints ». On a choisi le champ « Classified » comme « Target Field » et on a opté pour une stratégie d'échantillonnage « stratified random » avec 500 points aléatoires.

Ensuite, on a utilisé l'outil « Update Accuracy Assessment Points ». Dans l'« Input Raster or Feature Class Data », nous avons sélectionné la couche de buffer que l'on avait créée, unige CRS Buffers (une feature class), et dans l' « Input Accuracy Assessment Points », nous avons choisi la couche générée précédemment. Pour le « Target Field », nous avons sélectionné « Ground Truth ». Un message d'avertissement est apparu : « *Warning 001789: Not all points could be updated* ».

On a ensuite utilisé l'outil « Compute Confusion Matrix », en spécifiant dans « Input Accuracy Assessment Points » la couche « updated accuracy assessment points » qu'on venait de créer. On a de nouveau reçu un avertissement : « *Warning 001790: Not all points have valid class* ». On a répété cette procédure pour les images Landsat 8.

Afin d'évaluer la performance de la classification, plusieurs indicateurs sont calculés à partir de la matrice de confusion, parmi lesquels :

- Overall Accuracy (OA) : représente la proportion de pixels correctement classés, toutes catégories confondues. Il s'agit d'un indicateur global de performance.
- User Accuracy (UA) : pour chaque classe, il indique la proportion de pixels classés dans cette classe qui sont effectivement corrects. C'est une mesure de fiabilité pour l'utilisateur (d'où son nom).
- Producer Accuracy (PA) : c'est un outil essentiel qui permet d'évaluer la précision des processus de classification. Il calcule, pour chaque classe, le pourcentage de pixels qui ont été correctement identifiés par rapport à la réalité du terrain. Cette étape est essentielle pour garantir la complétude du modèle.
- Kappa coefficient : c'est un indicateur statistique qui évalue la concordance entre la classification établie et la réalité du terrain, en tenant compte de l'effet du hasard. Dans le cadre de notre analyse, il est essentiel de considérer la plage de variation de cet indicateur, qui s'étend de 0, représentant le hasard, à 1, indiquant un accord parfait.

Ces indicateurs permettent une évaluation détaillée de la qualité du modèle, en identifiant les classes bien ou mal reconnues.

Initialement, la matrice de confusion affichait uniquement des 0, et la valeur de Kappa était également de 0. Cela nous a conduit à reconsiderer la méthode. On a alors utilisé l'outil « Extract Values to Points » pour extraire les valeurs des points EarthTrack et les utiliser comme Ground Truth. Cela a ajouté une colonne « RASTERVALU » à la table attributaire.

Pour construire la matrice de confusion, on avait besoin de deux champs : « Ground Truth et Classified ». On a donc utilisé l'outil « Feature To Point » pour convertir la couche de buffer en points, puis on a appliqué « Extract Values to Points » en spécifiant la couche classifiée (Sentinel-2 ou Landsat 8) comme « Input Raster ». On a remarqué que les valeurs « RASTERVALU » allaient de 0 à 15, alors que les classes étaient numérotées de 1 à 16. On a donc ajusté manuellement les valeurs en ajoutant 1 à chaque « RASTERVALU » pour qu'elles correspondent à nos classes.

Enfin, on a renommé les champs « Classvalue » en « Classified » et « RASTERVALU » en « GrndTruth » pour qu'ils correspondent aux exigences de l'outil « Compute Confusion Matrix ». On a répété cette procédure pour les images Landsat 8.

## Layout

Au moment de créer le layout, on se rend compte que les valeurs de classification « Classvalue » ne correspondent pas aux valeurs utilisées dans le buffer. On doit donc tout vérifier manuellement. Cela est dû au fait qu'on avait utilisé la symbologie en fonction des commentaires « comments » pour la couche « unige\_CRS\_Buffers » et non des valeurs de classification « class\_value ». En revanche, pour la couche supervisée, on a utilisé la symbologie en fonction de la « class\_value ». On a utilisé le champ « comments » pour la couche buffer afin d'avoir les noms. Cependant, on a pu facilement adapter en consultant la table attributaire de chaque couche et en introduisant les bonnes valeurs manuellement pour les deux couches de classification supervisée, c'est-à-dire celles avec les satellites Landsat 8 et Sentinel-2.

Concernant le layout lui-même, on a veillé à intégrer une échelle, un titre, une rose des vents, ainsi que les légendes et les informations pertinentes pour les cartes.

Voici les catégories pour lesquelles on a choisi le code couleur, en s'inspirant de ce qui se fait généralement dans la cartographie de la couverture du sol, comme celles de CORINE Land Cover (*Sinergise, 2024*).

Tableau 1

N°	Catégorie	Couleur	Code Hex	Correspondance CORINE
1	Consolidated surface	Rouge foncé	#cc0000	122 - Road and rail networks
2	Buildings	Rouge vif	#e6004d	111 - Continuous urban fabric
3	Greenhouses	Rose magenta	#ff4dff	133 - Construction sites
4	Gardens	Jaune doré	#ffe64d	242 - Complex cultivation patterns
5	Lawns	Rose clair	#ffa6ff	141 - Green urban areas
6	Trees in artificial areas	Vert vif	#80ff00	311 - Broad-leaved forest
7	Mix of small structures	Violet clair	#cc4df2	121 - Industrial/commercial units
8	Grass & herbe vegetation	Vert citron	#ccf24d	321 - Natural grasslands
9	Short stem fruit trees	Orange clair	#f2a64d	222 - Fruit trees and berries
10	Vines	Orange foncé	#e68000	221 - Vineyards
11	Permanent garden plants & brush crops	Beige jaunâtre	#ffe6a6	241 - Annual crops with permanent crops
12	Closed forest	Vert foncé	#00a600	312 - Coniferous forest
13	Forest Strip	Vert néon	#4dff00	313 - Mixed forest
14	Linear woods	Vert anis	#a6f200	324 - Transitional woodland-shrub
15	Water	Bleu cyan	#00ccf2	511 - Water courses
16	Reedy marshes	Bleu lavande	#a6a6ff	411 - Inland marshes

Concernant les espaces urbains, le rouge et ses variantes ont été choisis pour représenter les surfaces consolidées et les bâtiments, conformément à la convention utilisée dans CORINE Land Cover (Sinergise, 2024).

Pour les zones agricoles et végétales, différentes nuances de vert et de jaune permettent de distinguer les cultures, les prairies et les formations arborées.

Pour les forêts et les espaces boisés, trois nuances de vert (foncé, clair et néon) ont été attribuées pour refléter la densité et le type de couvert forestier.

Pour les plans d'eau et les zones humides, le bleu cyan et le bleu lavande ont été utilisés pour les eaux de surface et les zones marécageuses, afin de faciliter leur reconnaissance.

Enfin, pour les structures particulières (serres, infrastructures industrielles, etc.), des couleurs spécifiques comme le rose magenta et le violet clair permettent d'identifier ces éléments distinctifs.

Le processus de classification supervisée a été réalisé en plusieurs étapes, comme illustré dans la Figure 4. Tout d'abord, les données ont été collectées sur le terrain à l'aide de l'application EarthTrack, permettant de géolocaliser et de classifier des points selon la nomenclature suisse NOLC04\_27. Ces données ont ensuite été importées dans ArcGIS Pro, où les bandes spectrales des images satellitaires (Sentinel-2 et Landsat 8) ont été composées pour créer des images en fausse couleur. Enfin, la classification supervisée a été effectuée en utilisant l'algorithme SVM (Support Vector Machine), avec une validation des résultats via une matrice de confusion.

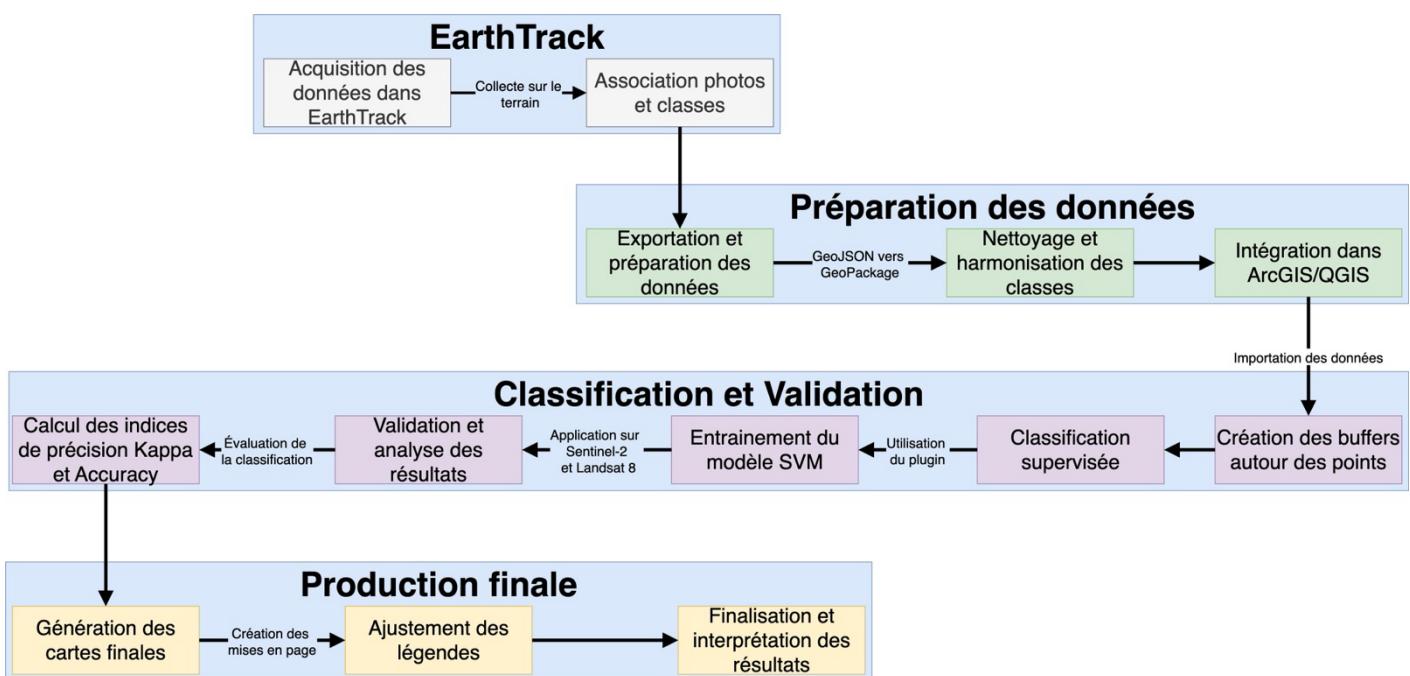


Figure 4 : Flux d'analyse pour la classification supervisée des images satellitaires (flowchart fait à l'aide de draw.io)

Comme le montre la Figure 4, le processus commence par la collecte de données sur le terrain via EarthTrack, où chaque point est géolocalisé et classifié selon la nomenclature NOLC04\_27. Ces points sont ensuite intégrés dans ArcGIS Pro, où les images satellitaires sont traitées pour créer des compositions spectrales. La classification supervisée est ensuite appliquée à l'aide de l'algorithme SVM, et les résultats sont validés à l'aide d'une matrice de confusion pour évaluer la précision de la classification.

## Présentation et analyse des résultats

Comme on peut le voir sur les figures en fausse couleur NIR 6 et 7 ci-dessous, les images satellitaires de Sentinel-2 sont beaucoup plus rouges que celles de Landsat 8, qui sont beaucoup plus claires. Les contrastes sont plus marqués. De plus, on remarque que l'eau du lac Léman est plus claire aux abords qu'en son centre, comme dans la figure 6, alors que c'est beaucoup plus homogène dans la figure 7. Concernant les classifications supervisées (figures 8 et 9), on remarque d'abord que les images à plus basse résolution distinguent l'eau en deux catégories distinctes (« forest\_strip » et « water »), alors que Sentinel-2 la catégorise comme une seule catégorie « water », ce qui était recherché.

C'est probablement la raison pour laquelle l'eau a été catégorisée en deux classes différentes lors de la classification supervisée utilisant l'image satellitaire Landsat 8, en raison des différentes couleurs de pixels de l'eau, alors que la classification supervisée via les images satellitaires Sentinel-2 ne retient qu'une seule et même catégorie, comme cela était prévu. La composition des bandes au format NIR montre ainsi une moins bonne absorption du proche infrarouge pour Landsat 8 que pour Sentinel-2. L'une des explications possibles de cette anomalie réside dans les caractéristiques spectrales des capteurs de Landsat 8.

À la différence de Sentinel-2, certaines bandes spectrales de Landsat offrent une capacité de pénétration minimale sous la surface de l'eau, notamment dans les zones peu profondes ou claires.

Une variation de la réflectance entre les bords du lac et son centre peut provoquer une classification erronée en deux classes distinctes.

Sentinel-2 semble moins affecté par ce phénomène, ce qui explique pourquoi il a mieux identifié le lac comme une seule entité homogène.

La densification des bâtiments dans la ville de Genève est beaucoup mieux distinguée sur la classification utilisant les bandes Sentinel-2 que sur celle utilisant les bandes Landsat 8, ce qui correspond davantage à la réalité. En effet, on voit une réelle zone, si l'on peut dire, « bétonnée » dans différentes catégories, alors qu'avec Landsat 8, les classifications mises en avant sur la carte sont beaucoup plus hétérogènes, c'est-à-dire qu'il y a beaucoup moins de catégories de type « bétonné », mais beaucoup plus de types de classification différents. On distingue donc moins bien la densification de la ville de Genève avec Landsat 8 qu'avec Sentinel-2.

La distinction entre les serres et les bâtiments est parfois également compliquée, surtout pour les serres les plus sophistiquées, car vues du ciel, elles ressemblent fortement à une toiture de bâtiment et les pixels sont donc très similaires.

En ce qui concerne les « closed\_forest », qui représentent les peuplements denses d'arbres, on constate que, pour les grandes forêts denses de Genève, les deux satellites donnent des résultats plutôt similaires.

En revanche, les deux satellites ont un peu plus de mal avec les autres types de végétation. En effet, les « forest\_strip » (bandes de forêts) ainsi que les « linear\_woods » (peuplements d'arbres linéaires) sont difficilement discernables par apprentissage automatique entre eux ainsi qu'avec les « closed\_forest ». Ce constat s'applique aux deux satellites.

On retrouve les mêmes problèmes de distinction des catégories de cultures horticoles permanentes et des arbres fruitiers à basse tiges, qui apparaissent très similaires sur les images satellites et ont pratiquement les mêmes pixels. Le gazon et la végétation herbacée sont également difficiles à distinguer, surtout selon le type de culture présent lors de la prise des images satellites, lesquelles peuvent montrer soit un sol nu (donc de couleur brun terreux), soit différentes teintes de vert (selon ce que le paysan décide de planter).

De même, les surfaces compactées (« consolidated\_surfaces ») sont pratiquement invisibles sur Landsat 8, alors qu'on en voit un peu plus avec Sentinel-2.

Les résultats de la matrice de confusion pour Sentinel-2 montrent une U\_Accuracy de 0,748503 et un Kappa de 0,730976. Pour Landsat 8, les résultats sont légèrement supérieurs, avec une U\_Accuracy de 0,820359 et un Kappa de 0,808134.

Contre toute attente, la classification basée sur les images Landsat 8 présente un meilleur score de Kappa, situé entre 0,8 et 1, ce qui correspond à un accord presque parfait selon l'échelle de Kappa du cours Geotool RS. En revanche, la classification Sentinel-2, bien que visuellement plus précise (notamment en évitant la double classe sur le lac Léman), obtient un accord fort mais inférieur à celui de Landsat 8. Cette différence sera approfondie dans la section suivante, celle de la discussion et de la conclusion.

$$K = (\text{Observations correctes} - \text{Observations attendues}) / (\text{Observations totales} - \text{Observations attendues})$$

Kappa est toujours  $\leq 1$  :

Accord très faible  $\leq 0.2$

Accord faible = 0.2 à 0.4

Accord modéré = 0.4 à 0.6

Accord fort = 0.6 à 0.8

Accord presque parfait = 0.8 à 1

Figure 5 : Échelle Kappa, provenant du cours geotool RS 2024

## Apport potentiel des données LiDAR pour affiner la classification

L'étude de Valence met en évidence l'intérêt d'intégrer des données LiDAR (*Light Detection and Ranging*) pour améliorer la classification de l'occupation du sol, en particulier dans les zones urbaines denses (Hermosilla et al., 2012).

Le LiDAR permet d'obtenir des informations en trois dimensions sur la hauteur des objets, ce qui est particulièrement utile pour distinguer des classes présentant des signatures spectrales similaires mais des structures différentes, comme :

- Bâtiments industriels vs immeubles résidentiels ;
- Forêts denses vs cultures basses ;
- Infrastructures urbaines vs terrains ouverts.

Dans le cadre de ce projet, l'ajout de données LiDAR pourrait permettre une meilleure distinction des classes en complétant les informations spectrales des images Sentinel-2 et Landsat 8. Une piste d'amélioration serait d'expérimenter un modèle numérique de surface (nDSM) pour intégrer cette dimension structurelle à la classification supervisée (Hermosilla et al., 2012).

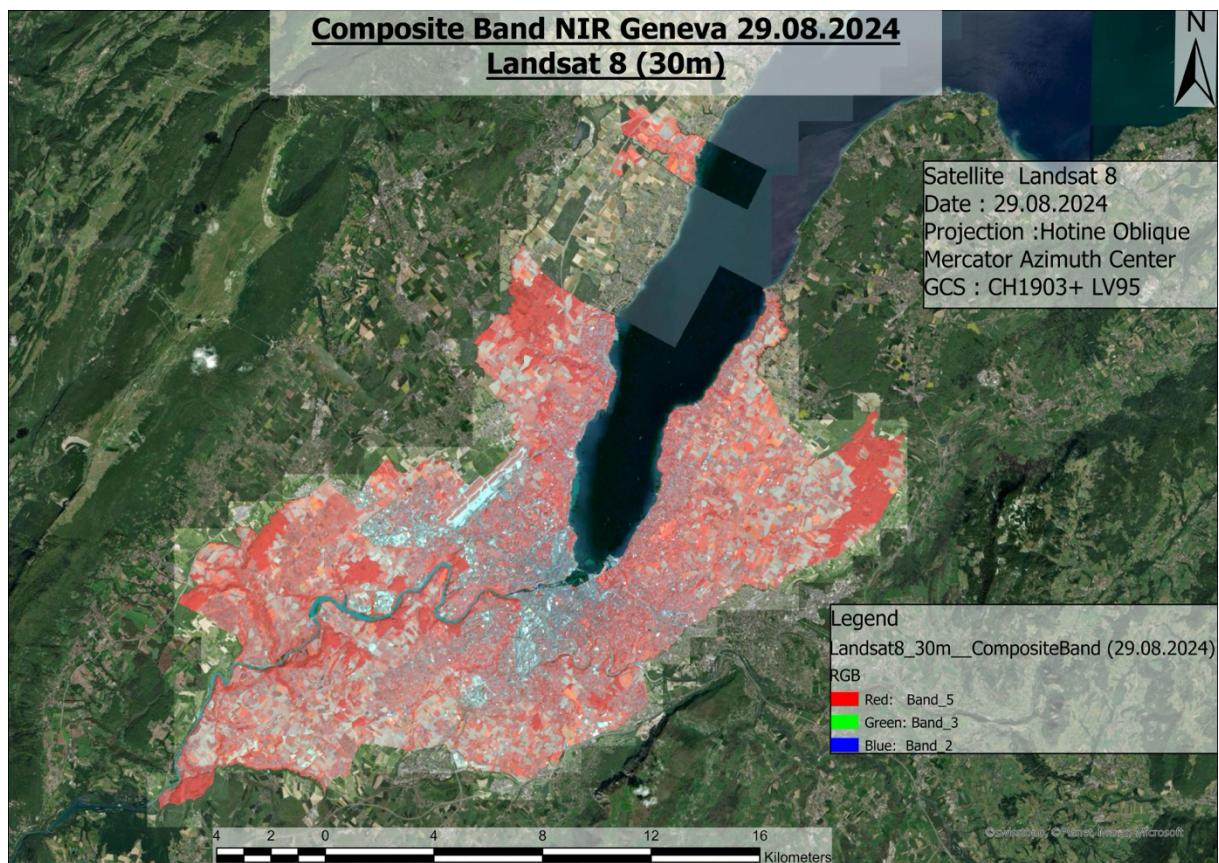


Figure 6 : Image satellitaire Landsat 8 en fausse couleur NIR montrant le lac Léman et Genève

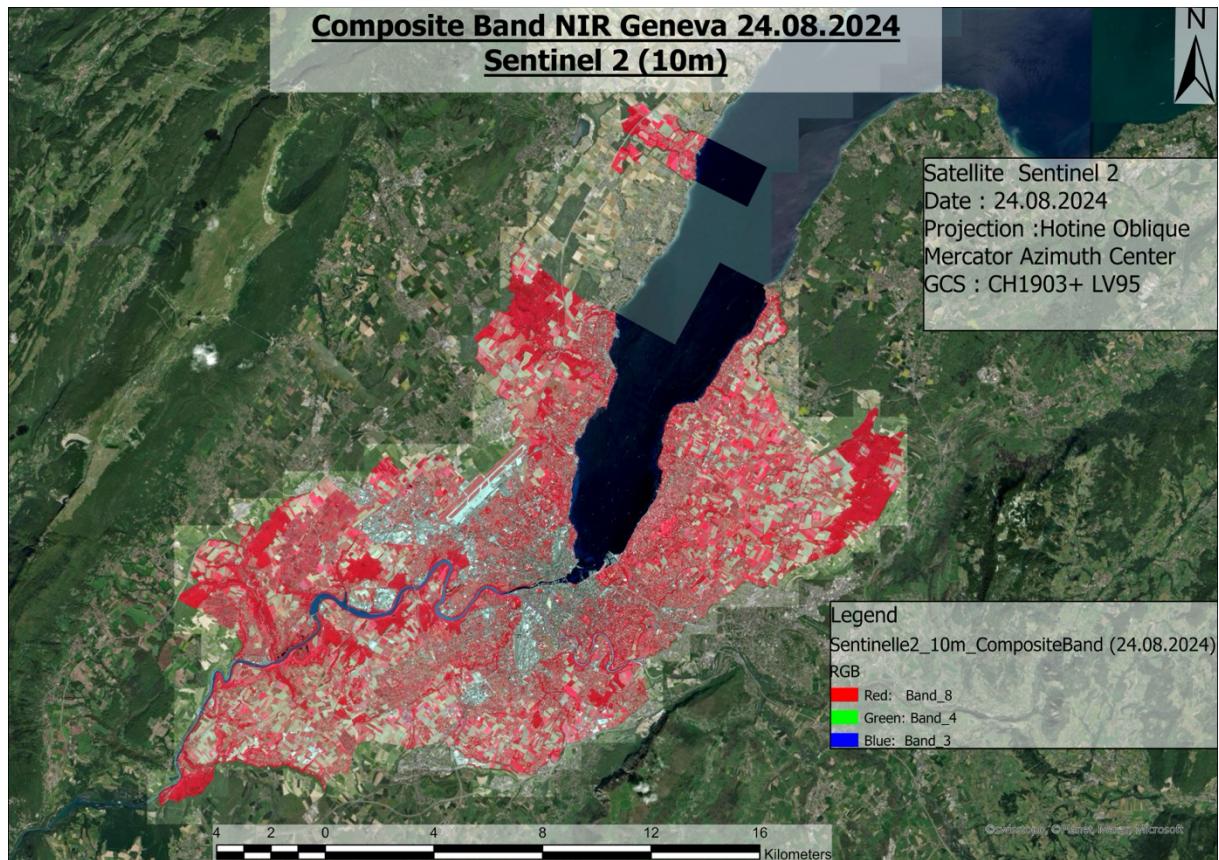


Figure 7 : Image satellitaire Sentinel-2 en fausse couleur NIR montrant le lac Léman et Genève

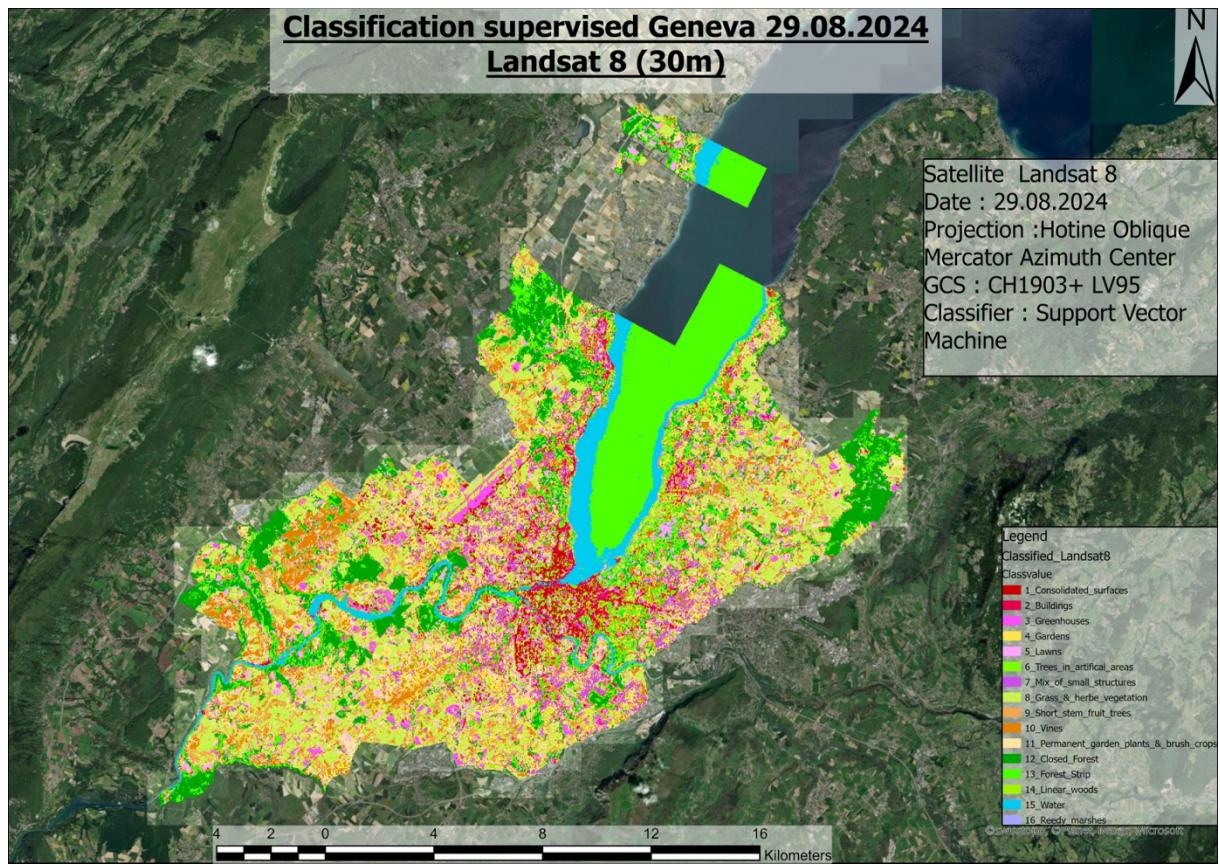


Figure 9 : Résultat de la classification supervisée utilisant les images Landsat 8

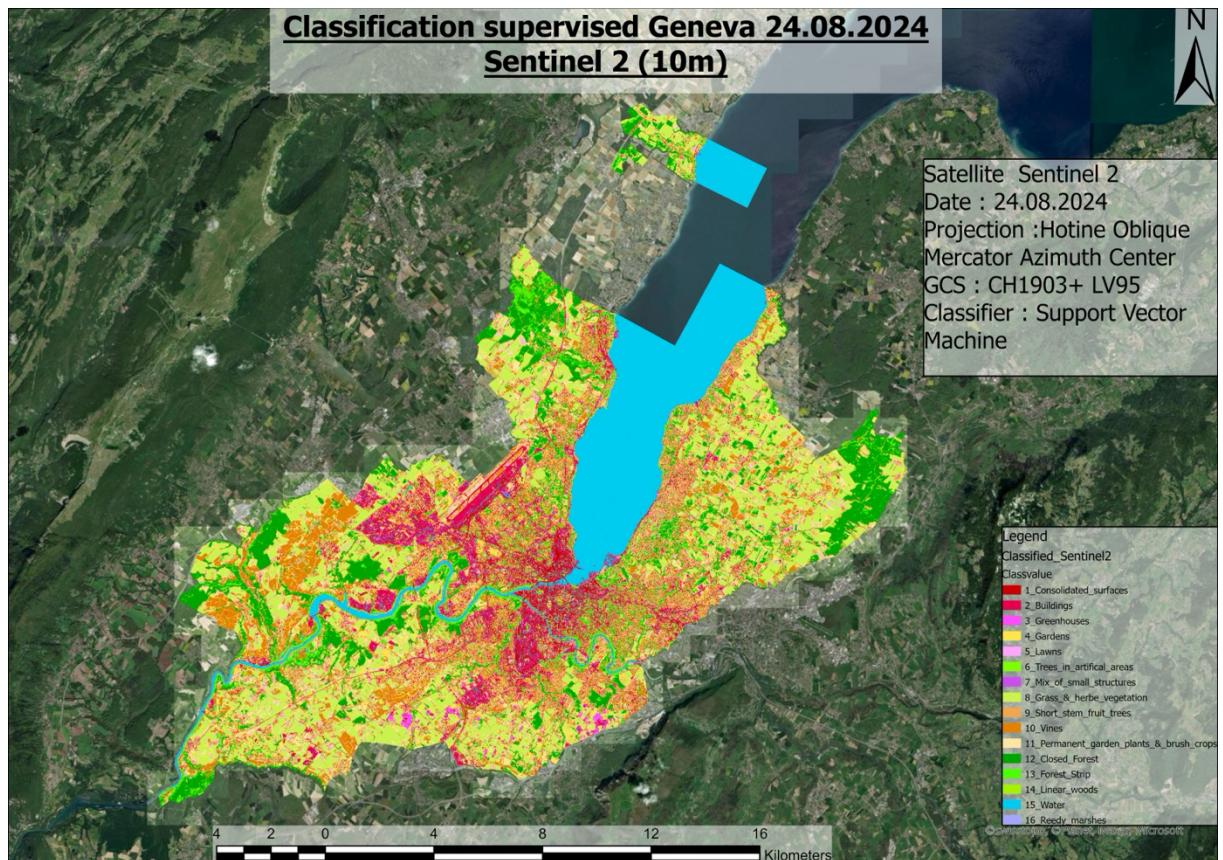


Figure 8 : Résultat de la classification supervisée utilisant les images Sentinel-2

## Discussion et Conclusions

Ce stage a permis d'explorer les potentialités et les limites de l'application EarthTrack dans le cadre d'une classification supervisée des images satellitaires. Grâce à ce stage, des compétences en télédétection, en analyse spatiale et en manipulation de données géographiques ont pu être acquises. Toutefois, plusieurs aspects méritent d'être discutés pour mieux évaluer la pertinence et l'efficacité de la méthodologie employée.

### **Limites de l'étude**

L'une des principales limites observées concerne la résolution des images satellites. Les images Landsat 8, avec une résolution de 30 mètres, se sont révélées moins adaptées que les images Sentinel-2, qui offrent une meilleure résolution (10 mètres), pour une classification fine du territoire. De plus, la distinction entre certaines classes de végétation s'est avérée difficile, ce qui suggère que la qualité des données d'entrée influence fortement la précision des résultats obtenus.

Une autre limite importante concerne le mode de validation du modèle.

Les points utilisés pour vérifier la qualité de la classification étaient les mêmes que ceux utilisés pour entraîner le modèle.

Théoriquement, cela devrait conduire à des taux de précision très élevés, voire parfaits, car le modèle est évalué sur des données qu'il connaît déjà.

Pourtant, les résultats montrent un score Kappa inférieur à 1, ce qui confirme que le modèle commet encore des erreurs.

Cette situation met en lumière les imperfections du modèle malgré l'auto-validation, et renforce l'idée que d'autres facteurs, comme la complexité spectrale des classes ou la confusion entre certaines catégories, peuvent affecter les performances, même avec des données connues.

Plus la résolution est élevée, plus les images permettent d'identifier des objets de plus en plus petits et d'obtenir des classifications de plus en plus précises. Alors qu'on s'attendait à ce que Sentinel-2, avec sa résolution plus fine, offre une meilleure précision de classification, les résultats montrent au contraire un score Kappa inférieur à celui de Landsat 8. Cette observation mérite d'être discutée et soulève plusieurs hypothèses.

Cette divergence entre les résultats visuels et les mesures quantitatives pourrait s'expliquer par la méthode de validation utilisée. En effet, seuls 167 points collectés via EarthTrack ont été utilisés pour évaluer la précision des classifications. Si un plus grand nombre de points, mieux répartis sur le territoire genevois, avait été utilisé, il est probable que la classification Sentinel-2 aurait obtenu de meilleurs scores, reflétant davantage la précision visuelle observée. 167 est un nombre assez limité pour construire une évaluation solide, surtout pour 16 classes. Cette faible densité peut affecter plus négativement Sentinel-2, qui détecte plus de détails.

De plus, la répartition spatiale des points de vérification a peut-être introduit un biais dans les résultats. Par exemple, dans la classification issue des images Landsat 8, deux classes distinctes apparaissent dans le Léman, alors que la classification Sentinel-2 regroupe correctement l'ensemble du lac sous une seule classe, ce qui constituait l'objectif recherché.

Si les points d'évaluation générés aléatoirement (via la méthode stratifiée aléatoire) n'ont pas couvert de manière suffisante les zones problématiques du Léman, ou si les points sont tombés uniquement dans l'une des deux classes créées à tort, cela pourrait avoir conduit à une surestimation de la précision de Landsat 8.

Une autre limite importante concerne la contrainte physique d'accès lors de la collecte des données avec EarthTrack. Certaines zones, comme le milieu du lac Léman, des parcelles privées ou des milieux naturels difficiles d'accès, n'ont pas pu être atteintes. Cela empêche une sélection totalement aléatoire des points, comme le voudrait une approche statistique rigoureuse. Ce biais d'accès peut conduire à une surreprésentation des milieux urbains ou faciles d'accès, et à une sous-représentation d'autres classes importantes pour la classification.

Concernant la faisabilité du projet à grande échelle, l'utilisation d'EarthTrack s'est avérée efficace pour la collecte de points géolocalisés à Genève, mais l'application du même processus à l'échelle nationale soulève plusieurs problèmes. La collecte manuelle de points sur le terrain est lente et nécessiterait une communauté active de contributeurs pour être étendue à l'ensemble du territoire. Sans cette communauté active, l'approche pourrait ne pas être viable pour une cartographie systématique du territoire suisse, et a fortiori, pas ailleurs.

## Efficacité d'EarthTrack

L'application EarthTrack peut s'avérer utile dans certains cas, notamment pour identifier des éléments cachés sous la canopée, tels que des routes ou des cours d'eau recouverts par la végétation. Cependant, il existe une limite à cela : le pixel correspondrait à de la végétation, ce qui augmenterait la probabilité d'erreur dans l'interprétation du modèle de classification. Dans ce cas précis, EarthTrack serait particulièrement précieux s'il était accompagné d'images, permettant une correction manuelle dans un SIG à l'aide de polygones. Aujourd'hui, pour contourner ce problème, on utilise par exemple des images satellitaires en hiver, dans les régions où les arbres perdent leurs feuilles durant cette saison.

Une approche basée uniquement sur des polygones dessinés directement dans un SIG pourrait parfois s'avérer plus rapide et tout aussi précise. L'apport d'EarthTrack doit donc être évalué en fonction des spécificités du terrain et des objectifs du projet.

Au-delà de l'aspect technique, les résultats de cette étude présentent un réel potentiel d'application dans des domaines variés. En aménagement urbain, une classification plus précise permettrait d'identifier les zones densément bâties, les espaces verts ou les surfaces imperméables, facilitant ainsi la planification territoriale. En écologie, la capacité à distinguer les types de forêts, bandes boisées ou zones humides est essentielle pour le suivi des milieux naturels, bien que cela reste encore difficile à réaliser uniquement à partir d'images satellitaires. Enfin, en agriculture, l'intégration d'indices spectraux comme le NDVI (Normalized Difference Vegetation Index) permettrait de suivre l'état des cultures ou de détecter des zones de stress hydrique. Ce type d'approche est particulièrement utile pour les politiques de gestion durable des ressources agricoles.

Il est également important de noter qu'il existe de nombreux autres indices spectraux qui peuvent améliorer la précision des classifications. Des indices comme le NDWI

(Normalized Difference Water Index) pour détecter les zones humides, le NDBI (Normalized Difference Built-up Index) pour les surfaces bâties, ou encore le GCI (Green Chlorophyll Index) pour estimer la biomasse végétale, sont autant d'outils disponibles. Dans le cadre de ce projet, ces indices n'ont pas été intégrés, ce qui représente une limite méthodologique. Leur utilisation combinée aurait pu permettre de mieux discriminer certaines classes visuellement proches, notamment dans les zones agricoles ou mixtes. Cela constitue une piste d'amélioration importante pour de futures études.

En résumé, l'utilisation d'EarthTrack a facilité la collecte de données terrain grâce à son interface intuitive, sa géolocalisation automatique et l'enregistrement de photos associées à chaque point. Cet outil s'est révélé efficace pour construire rapidement un jeu d'entraînement compatible avec un SIG.

Toutefois, plusieurs limites sont à relever. Premièrement, la nomenclature utilisée par défaut (FAO LCCS) n'est pas adaptée au contexte suisse, ce qui a nécessité un travail de correspondance avec la NOLC04\_27. Deuxièmement, la collecte de points est fortement contrainte par l'accessibilité physique : il n'est pas possible d'aller collecter des données au milieu du lac Léman, dans des propriétés privées ou des zones naturelles protégées.

Cela empêche une sélection véritablement aléatoire, ce qui introduit un biais spatial en faveur des zones faciles d'accès (zones urbaines, routes, parcs). Enfin, étant seul à contribuer à cette base de données, la densité des points reste limitée, ce qui peut réduire la représentativité de certaines classes.

Pour surmonter ces limites, une piste d'amélioration serait de développer une communauté active d'utilisateurs d'EarthTrack, ou de croiser ces données avec d'autres sources (digitalisation SIG, LiDAR, observations ciblées).

## Recommandations et perspectives d'avenir

Pour améliorer la précision des classifications futures, plusieurs recommandations peuvent être envisagées :

- Explorer d'autres sources de données satellitaires offrant une meilleure résolution spatiale et spectrale.
- Augmenter le nombre de points de terrain afin d'améliorer l'entraînement du modèle de classification.
- Combiner plusieurs approches : un croisement entre une collecte manuelle et une digitalisation assistée dans un SIG pourrait offrir un bon compromis entre précision et efficacité.
- Développer un projet pilote à plus grande échelle, en testant l'application EarthTrack sur un territoire plus vaste, avec une communauté d'utilisateurs plus large.

Enfin, ce travail marque le début d'une nouvelle série de recherches et d'expérimentations. Une autre étude intéressante pourrait consister à utiliser EarthTrack pour affiner les classifications en ciblant les zones où les modèles d'apprentissage automatique présentent des incertitudes. Les chercheurs pourraient identifier ces zones et inviter la communauté EarthTrack à y collecter des points géolocalisés

accompagnés de photos. Ces informations permettraient de valider ou de corriger les classifications et d'affiner les ajustements dans les SIG.

Une autre approche complémentaire pourrait consister à combiner les observations satellites et les relevés sur le terrain pour affiner les classifications. Par exemple, si une zone est mal identifiée sur les images satellites, les chercheurs pourraient solliciter des contributions ciblées via EarthTrack afin de confirmer la nature du sol. Cette méthode permettrait d'améliorer la gestion des terres et de réduire les erreurs d'interprétation, comme la confusion entre une forêt dense et une simple ombre sur une image satellite. Ce stage a donc constitué une étape enrichissante et formatrice, renforçant mon intérêt pour la télédétection et la géomatique. Je suis désormais mieux armé pour aborder des problématiques similaires et contribuer à des projets de cartographie et d'analyse spatiale de manière plus approfondie.

## Réflexions sur le déroulement du stage

Le stage s'est déroulé sur une période de 12 mois par intermittence, principalement à l'Université de Genève, avec des sorties sur le terrain pour la collecte de données. J'ai travaillé en étroite collaboration avec M. Chognard, qui m'a fourni un soutien technique, ainsi qu'avec M. Giuliani, qui m'a apporté un soutien méthodologique tout au long du projet. Les ressources matérielles (ordinateurs, logiciels) étaient suffisantes, même si j'ai rencontré des difficultés techniques avec QGIS, principalement en raison de la configuration de mon ordinateur personnel.

Ce stage m'a permis de développer mes compétences en télédétection, en classification supervisée et en analyse spatiale. J'ai également appris à utiliser des outils comme ArcGIS Pro et à adapter des applications existantes (EarthTrack) à des besoins spécifiques. Cependant, j'ai dû approfondir mes connaissances en Python pour résoudre certains problèmes techniques, ce qui a été une expérience enrichissante. Je recommanderais ce stage à des étudiants intéressés par ces domaines. Il est essentiel d'avoir de bonnes connaissances en SIG et en analyse de données avant de commencer, ainsi qu'une certaine autonomie pour résoudre des problèmes techniques. Une bonne communication avec les superviseurs est également indispensable pour progresser efficacement.

## Annexe:

### Définition EarthTrack:

#### **Cheminement:**

**Landcover --> not vegetated --> Natural --> Natural Bare Surfaces --> type of natural bare surface:**

#### **Bare rocks :**

Roches nues : la surface de la roche est continue, sauf peut-être pour quelques fissures dans le matériau. Certaines zones peuvent être recouvertes de couches superficielles de sol ou il pourrait y avoir des poches isolées de sol ou un mélange des deux.

#### **Gravel stones and/or boulders:**

Pierres de gravier et/ou gros rochers : Au moins 60% du total des fragments grossiers sont constitués de l'un des trois.

#### **Bare soil, mud or hard sand:**

Sol nu, boue ou sable dur : matériaux non consolidés résultant de l'altération du matériau parent (y compris les effets de l'humidité et de la température) et/ou des macro- et micro-organismes.

#### **Loose and shifting sands :**

Sables lâches et mouvants : zones recouvertes de particules de sol. Ces particules peuvent être déplacées par des vents réguliers et former des motifs distincts. Une surface pierreuse peut être spécifiée.

#### **Gilgai :**

Gilgai : micro-relief typique des vertisols (teneur élevée en minéraux argileux), qui se dilatent et se contractent largement avec des changements saisonniers distincts de la teneur en humidité

#### **Cheminement:**

**Landcover --> not vegetated --> human-made surface --> type of human-made surface:**

#### **Paved road or surface :**

Route ou surface pavée : recouverte d'un matériau artificiel pour consolider la surface du sol.

#### **Urban areas (non linear) :**

Zones urbaines (non linéaires) : zones construites non linéaires recouvertes de structures imperméables adjacentes à ou connectées par des rues.

### **Industrial and/or other areas:**

Zones industrielles et/ou autres : Surfaces imperméables non linéaires liées au commerce, à la fabrication, à la distribution et au commerce (par exemple, aéroports, ports, usines)

### **Unpaved road or surface :**

Route ou surface non pavée : soit une barre et consiste en une surface comprimée ou recouverte de matériel non consolidé (par exemple, du gravier)

### **Waste dump :**

Décharge : où le terrain existant ou la couverture terrestre est recouvert de matériaux provenant d'une source extérieure (artefacts et matériaux transportés par les humains)

### **Ploughed or tilled fields :**

Champs labourés ou cultivés : zones de terre cultivée qui sont temporairement sans couverture végétale en raison du labour ou du travail du sol

### **Cheminement:**

### **Landcover --> vegetated --> no Trees --> Not Cultivated --> life form:**

#### **Trees :**

Arbres : plantes pérennes ligneuses avec une (bien que pouvant être multiple) tige bien définie portant des couronnes plus ou moins définies, d'au moins 1m de haut ou potentiellement capables d'atteindre in situ. Peut atteindre jusqu'à 3-5m de hauteur. Comprend les bambous.

#### **Tall shrubs :**

Grands arbustes : plantes pérennes ligneuses avec des tiges ligneuses persistantes et sans tige principale définie, mesurant moins de 5m de haut. L'habitude de croissance peut être dressée, étalée ou prostrée.

#### **Dwarf shrubs :**

Arbustes nains : arbustes qui mesurent 1m ou moins (par exemple, les ajoncs, les bruyères et autres éricoïdes).

#### **Forbs :**

Herbes : plantes herbacées à feuilles larges au sens commun (par exemple, les tournesols). Comprend toutes les plantes herbacées non graminées, les fougères (sauf les fougères arborescentes) et les succulentes très basses et non feuillues.

#### **Ferns and allies :**

Fougères et alliées

#### **Grasses, rushes and/or sedges:**

Graminées, joncs et/ou carex : toutes les graminées, joncs et carex herbacés. Exclut les formes ligneuses comme les bambous.

### **Mosses :**

Mousses : plantes terrestres photoautotrophes sans vraies feuilles, tiges ou racines, mais avec des organes ressemblant à des feuilles et des tiges (par exemple, les espèces de Sphagnum).

### **Lichens :**

Lichens : organismes composites formés par l'association symbiotique de champignons et d'algues. On les trouve incrustés sur les rochers, les troncs d'arbres, etc.

### **Cheminement:**

**Landcover --> vegetated --> no Trees --> Cultivated --> life form:**

#### **Graminoids :**

Graminoïdes : comprend toutes les graminées ainsi que d'autres plantes ressemblant à des graminées à feuilles étroites qui ne sont pas des graminées selon la taxonomie.

#### **Non-graminoids :**

Non-graminoïdes : toutes les plantes herbacées qui ne sont pas des gramoïdes (c'est-à-dire, les espèces qui n'appartiennent pas aux graminées selon la définition taxonomique). Peut couvrir de grandes surfaces d'eau si aquatique.

#### **Shrubs :**

Arbustes : plantes pérennes ligneuses avec des tiges persistantes et ligneuses et sans tige principale définie. L'habitude de croissance peut être dressée, étalée ou prostrée. Les plantes forment souvent un bloc distinct et sont souvent plantées à un espacement ou un motif régulier (par exemple, les vergers ; les pépinières).

## Tableau EarthTrack x NOLC04\_27

Tableau 2

NOLC04_6	NOLC04_27	Cheminement Dans EarthTrack	Explication
10. Artificial areas	11 - Consolidated surfaces	<b>Landcover --&gt; not vegetated --&gt; human-made surface --&gt; type of human-made surface --&gt; Paved Road or surface”</b>	“Paved road or surface” qui est défini comme une surface recouverte d'un matériau artificiel pour consolider la surface du sol. Cette catégorie englobe les surfaces non naturelles, partiellement ou totalement imperméabilisées, y compris les sols asphaltés, bétonnés, marneux, couverts de gravier, de pavés ou de dalles. Elle comprend également les surfaces partiellement ou totalement imperméabilisées par des mesures de construction, les sols naturels fortement tassés, les rues, chemins, trottoirs, places, ponts, voies ferrées jusqu'à la transition avec une autre couverture, les pistes en dur des aérodromes, les abris isolés aux arrêts et haltes de transports publics, les aires de stockage du fumier, réservoirs de lisier à ciel ouvert, silos-tranchées, les bassins en béton de piscines, de parcs, de stations d'épuration, de bassins de pisciculture, les bassins de compensation en béton de centrales électriques, plans d'eau pour la lutte contre les incendies, les surfaces compactées de chantiers et les surfaces compactées sous des arbres.
	12 - Buildings	<b>Landcover --&gt; not vegetated --&gt; human-made</b>	“Urban areas (non linear)” qui est défini comme des zones construites non linéaires

		<p><b>surface --&gt; type of human-made surface -&gt; Urban areas (nonlinear)</b></p>	<p>recouvertes de structures imperméables adjacentes à ou connectées par des rues. Cette catégorie englobe les constructions à caractère permanent, recouvertes d'un toit, destinées à abriter, loger ou placer des personnes, des animaux ou du matériel. Elle comprend également les maisons individuelles et maisons de deux logements, immeubles d'habitation, chalets, maisons de vacances, maisons en terrasses y compris les toits végétalisés, les garages, les pavillons et les jardins d'hiver, les fermes, granges, remises, étables dans les champs, habitations d'exploitations agricoles, « Stöckli », poulaillers et ruchers, fenils, mazots, cabanes d'alpage, les bâtiments industriels, artisanaux et commerciaux, y compris les halles de production et de stockage, les centres commerciaux, les magasins de meubles, les kiosques, les banques, bâtiments administratifs, églises, gares, restaurants, les bâtiments destinés aux activités de loisirs, cabanes de clubs, halles de sport, vestiaires de terrains de sport, les tribunes couvertes de stades, les ruines si le point d'échantillonnage se trouve sur les restes du bâtiment (fondations, combles), les silos et réservoirs, le bâtiment visible en forêt, les caravanes sur un emplacement aménagé comme un terrain attenant à une maison individuelle (jardin, chemins, places fixes pour s'asseoir, etc.),</p>
--	--	---	---

			les toits plats végétalisés, et les bâtiments en construction, dont la toiture est achevée.
13 - Greenhouses	<b>Landcover --&gt; not vegetated --&gt; human-made surface --&gt; type of human-made surface -&gt; Industrial and/or other areas</b>	"Industrial and/or other areas" qui est défini comme des surfaces imperméables non linéaires liées au commerce, à la fabrication, à la distribution et au commerce (par exemple, aéroports, ports, usines). Cette catégorie englobe les constructions en verre ou tunnels en plastique, à caractère permanent, destinées à la culture de plantes. Elle comprend également les serres avec ou sans fondations, les cultures hors-sols, les tunnels en plastique avec ou sans fondations, à hauteur d'homme, les tunnels en plastique pour mettre le matériel et les machines agricoles à l'abri (Agribuchs).	
14- Gardens	<b>Landcover --&gt; vegetated --&gt; no Trees --&gt; Cultivated --&gt; life form --&gt; non-graminoids</b>	"Non-graminoids" qui est défini comme toutes les plantes herbacées qui ne sont pas des graminoides (c'est-à-dire, les espèces qui n'appartiennent pas aux graminées selon la définition taxonomique). Cette catégorie englobe les surfaces morcelées par des planches et de petites allées, plantées d'espèces végétales diverses. Elle comprend également les châssis pour cultures forcées, les surfaces de jardineries, d'exploitations horticoles et de pépinières, morcelées par de petites allées et plantées d'espèces végétales diverses, les planches de légumes, d'herbes aromatiques et de fleurs sur les terrains attenants	

			aux bâtiments, les petites allées entre les planches, les planches de légumes, d'herbes aromatiques et de fleurs dans les jardins familiaux, les parcelles jardinées sur des terres arables, les petites parcelles plantées de fleurs diverses sur des terres arables, pour l'autocueillette, les rangées de tombes dans les cimetières, les parterres de fleurs dans les parcs publics et les terrains d'exposition, les arbres sur des structures jardinées (selon la liste des priorités), et les petites cabanes à outils dans les jardins familiaux et sur les terrains attenants aux bâtiments, entourées par des structures jardinées.
15 - Lawns	<b>Landcover --&gt; vegetated --&gt; no Trees --&gt; Cultivated --&gt; life form --&gt; Graminoids</b>	"Graminoids" qui est défini comme comprenant toutes les graminées ainsi que d'autres plantes ressemblant à des graminées à feuilles étroites qui ne sont pas des graminées selon la taxonomie. Cette catégorie englobe les surfaces herbeuses homogènes, pauvres en espèces, entretenuées de manière intensive et sans affectation agricole. Elle comprend également le gazon de terrain de sport et de terrain de golf, le gazon ornemental, la végétation des endroits piétinés sur les terrains attenants aux bâtiments, le gazon d'exploitation horticole, les terrains de jeux engazonnés des parcs publics, les pelouses des piscines, les îlots directionnels engazonnés, et les toits	

		engazonnés de garages souterrains.
16 - Trees in artificial areas	<b>Landcover -&gt; vegetated -&gt; trees-&gt; tree crop or urban trees --&gt; trees in artificial areas</b>	<p>“trees in artificial areas” car elle est spécifiquement conçue pour les arbres qui sont cultivés (tree crop) ou qui se trouvent dans des zones urbaines (urban trees).</p> <p>La description que vous avez donnée pour “trees in artificial areas” mentionne des arbres situés dans diverses zones artificielles, y compris des terrains attenants aux bâtiments, des terrains de camping, des cimetières et des terrains de sport. Ces zones sont généralement considérées comme des zones urbaines, ce qui rend la catégorie “urban trees” appropriée.</p> <p>De plus, la mention des “3 arbres (fruitiers) de plus de 3m de haut et distants de 25 m au maximum (distance entre les troncs) sur du gazon ou de petites cultures mixtes” pourrait être interprétée comme une forme d’arboriculture, ce qui correspond à la catégorie “tree crop”.</p>
17 - Mix of small structures	<b>Landcover --&gt; vegetated --&gt; no Trees --&gt; Not Cultivated --&gt; life form:</b> <ul style="list-style-type: none"> <li>- Trees</li> <li>- Tall shrubs</li> <li>- Forbs</li> <li>- Grasses, rushes and/or sedges</li> </ul>	<p><b>Landcover --&gt; vegetated --&gt; no Trees --&gt; Not Cultivated --&gt; life form”.</b></p> <p>En effet, ces espaces alternent sur de petites surfaces du gazon, des arbustes, des arbres isolés et des fleurs. Ils peuvent donc être classés dans plusieurs sous-catégories en fonction de la végétation dominante :</p> <ol style="list-style-type: none"> <li>1. <b>“Trees”</b> : Si les arbres isolés sont la</li> </ol>

			<p>caractéristique dominante.</p> <ol style="list-style-type: none"> <li>2. <b>"Tall shrubs"</b> : Si les arbustes sont la caractéristique dominante et mesurent moins de 5 mètres de haut.</li> <li>3. <b>"Forbs"</b> : Si les fleurs et autres plantes herbacées à feuilles larges sont la caractéristique dominante.</li> <li>4. <b>"Grasses, rushes and/or sedges"</b> : Si le gazon est la caractéristique dominante.</li> </ol>
20 - Grass & herb vegetation	21 - Grass & herb vegetation	<b>Landcover --&gt; vegetated --&gt; no trees --&gt; Not cultivated --&gt; grass &amp; herb vegetation</b>	<p>La catégorie "Forbs" a été choisie pour "grass &amp; herb vegetation" car elle englobe les plantes herbacées à feuilles larges, ce qui pourrait inclure une variété d'herbes et de plantes mentionnées dans la description. Cependant, le choix dépend de la composition spécifique de la végétation dans les zones à classer. Si elles sont principalement des plantes à feuilles larges, alors "<b>Forbs</b>" est approprié. Si elles sont principalement des graminées, une autre catégorie pourrait être plus appropriée, comme "<b>Grasses, rushes and/or sedges</b>".</p>
30 - Brush vegetation	31 - Shrubs	<b>Landcover --&gt; vegetated --&gt; no Trees --&gt; Not Cultivated --&gt; Tall shrubs</b>	<p>Cette catégorie englobe les plantes pérennes ligneuses avec des tiges ligneuses persistantes et sans tige principale définie, mesurant moins de 5m de haut. Cela correspond à la description de "shrubs" qui comprend des</p>

			surfaces couvertes de buissons, d'arbustes ou d'arbustes nains dont la hauteur dominante ne dépasse pas 3 m. De plus, cette catégorie pourrait également inclure des surfaces boisées d'arbres d'une hauteur dominante inférieure à 3 m et des aires afforestées dont la hauteur dominante des arbres est inférieure à 3 m, qui sont également mentionnées dans la description de "shrubs".
32-Bush meadows	<b>Landcover --&gt; vegetated --&gt; no Trees --&gt; Not Cultivated --&gt; life form: Shrubs</b>	Cette catégorie englobe les plantes pérennes ligneuses avec des tiges ligneuses persistantes et sans tige principale définie. Cela correspond à la description de "brush meadows" qui comprend des surfaces couvertes principalement de buissons, d'arbuste et d'arbustes nains dont la hauteur dominante ne dépasse pas 3 m. De plus, cette catégorie pourrait également inclure des surfaces boisées d'arbres d'une hauteur dominante inférieure à 3 m et des aires afforestées dont le degré de couvert est de 50-80% et la hauteur dominante inférieure à 3 m, qui sont également mentionnées dans la description de "brush meadows".	
33-Short-stem fruit trees	<b>Landcover --&gt; vegetated --&gt; Trees --&gt; Tree Crop or Urban Trees</b>	Cette catégorie englobe les arbres établis par plantation et/ou semis délibéré (par exemple, forêt plantée ou de plantation, vergers d'arbres, agroforesterie) ou des arbres dans des contextes urbains (par exemple, dans des parcs ou des jardins ou le long des routes).	

			Cela correspond à la description de “short-stem fruit trees” qui comprend des surfaces plantées d’arbres fruitiers à basses tiges et mitiges. De plus, cette catégorie pourrait également inclure des plantations d’arbres fruitiers (fruits à noyau ou à pépins), champs d’abricotiers, plantations d’arbres à kiwis comptant au moins 3 arbres par ares, qui sont également mentionnés dans la description de “short-stem fruit trees”.
34 - Vines	<b>Landcover --&gt; vegetated --&gt; no Trees --&gt; Not Cultivated --&gt; life form: Shrubs</b>	Cette catégorie englobe les plantes pérennes ligneuses avec des tiges persistantes et ligneuses et sans tige principale définie. Cela correspond à la description de “vines” qui comprend des surfaces plantées de vignes. De plus, cette catégorie pourrait également inclure des plantations nouvelles ou renouvelées, des vignes en pergola, et des pépinières de vignes, qui sont également mentionnées dans la description de “vines”.	
35 - Permanent garden plants & brush crops	<b>Landcover --&gt; vegetated --&gt; no Trees --&gt; Cultivated --&gt; life form: Shrubs</b>	Cette catégorie englobe les plantes pérennes ligneuses avec des tiges persistantes et ligneuses et sans tige principale définie. Cela correspond à la description de “permanent garden plants & brush corps” qui comprend des surfaces plantées de cultures de petits fruits, de houblon, de plantes médicinales et aromatiques résistantes, de plantes vivaces, d’arbustes et d’arbres dans des pépinières et surfaces plantées	

			de sapins de Noël. De plus, cette catégorie pourrait également inclure des arbres et arbustes dans des pépinières, des arbres dans des pépinières et pépinières en forêt et à l'extérieur de la forêt, des plantations d'arbres de Noël, et des plantations pluriannuelles de petits fruits, qui sont également mentionnés dans la description de “permanent garden plants & brush corps”.
40-Tree vegetation	41 - Closed forest	<b>Landcover --&gt; vegetated --&gt; Trees --&gt; Forest or Woodland</b>	Cette catégorie englobe les zones de végétation naturelle semi-naturelle, avec des arbres comme forme de vie dominante. Cela correspond à la description de “closed forest” qui comprend des peuplements d'arbres de plus de 50 m de long et de large et peuplements de 50 m de long au maximum et de 25- 50 m de large qui les jouxtent, dont la hauteur dominante dépasse 3 m, et le degré de couvert à l'intérieur de la surface de référence est supérieur à 60%. De plus, cette catégorie pourrait également inclure des peuplements d'espèces forestières et d'espèces communes aux parcs, des arbres fruitiers à hautes tiges (fruits à pépins ou à noyau, noyers), des plantations de peupliers et de saules, et des aires afforestées dont la hauteur dominante est supérieure à 3 m, qui sont également mentionnés dans la description de “closed forest”.
	42-Forest edges	<b>Landcover --&gt; vegetated --&gt;</b>	Cette catégorie englobe les zones de végétation naturelle semi-naturelle, avec des arbres

	<b>Trees --&gt; Forest or Woodland</b>	comme forme de vie dominante. Cela correspond à la description de “forest edges” qui comprend des peuplements d’arbres de moins de 25 m de long et de large, jouxtant un peuplement de plus de 25 m de large, dont la hauteur dominante dépasse 3 m, et le degré de couvert à l’intérieur de la surface de référence est supérieur à 60%. De plus, cette catégorie pourrait également inclure des peuplements d’espèces forestières et d’espèces communes aux parcs, des arbres fruitiers à hautes tiges (fruits à pépins ou à noyau, noyers), des plantations de peupliers et de saules, et des aires afforestées dont la hauteur dominante est supérieure à 3 m, qui sont également mentionnés dans la description de “forest edges”.
43-Forest strips	<b>Landcover --&gt; vegetated --&gt; Trees --&gt; Forest or Woodland</b>	Cette catégorie englobe les zones de végétation naturelle semi-naturelle, avec des arbres comme forme de vie dominante. Cela correspond à la description de “forest strips” qui comprend des peuplements d’arbres isolés et des bandes de forêt de 25-50 m de large, à condition que le point d’échantillonnage soit à l’intérieur de la lisière du boisement, la hauteur dominante dépasse 3 m, le degré de couvert à l’intérieur de la surface de référence soit supérieur à 60% et que la part de forêt buissonnante représente moins des 2/3. De plus, cette catégorie pourrait également inclure des peuplements d’espèces

		forestières et d'espèces communes aux parcs, des arbres fruitiers à hautes tiges (fruits à pépins ou à noyau, noyers), des plantations de peupliers et de saules, des aires afforestées dont la hauteur dominante est supérieure à 3 m, et des cours d'eau de moins de 6 m de large coulant dans une bande de forêt, qui sont également mentionnés dans la description de "forest strips".
44 - Open forest	<b>Landcover --&gt; vegetated --&gt; Trees --&gt; Forest or Woodland</b>	<p>Cette catégorie englobe les zones de végétation naturelle semi-naturelle, avec des arbres comme forme de vie dominante. Cela correspond à la description de "open forest" qui comprend des peuplements d'arbres de 50 m de large au minimum, espacés temporairement ou de manière permanente, à condition que le point d'échantillonnage soit à l'intérieur de la lisière du boisement, la hauteur dominante dépasse 3 m, le degré de couvert de la surface de référence soit compris entre 20 et 60% et que la part de forêt buissonnante représente moins des 2/3.</p> <p>De plus, cette catégorie pourrait également inclure des peuplements d'arbres clairsemés sur des surfaces agricoles en plaine et en montagne, des peuplements d'arbres clairsemés sur des surfaces improductives (pas utilisées à des fins agricoles), des peuplements d'espèces forestières et d'espèces communes aux parcs, des arbres fruitiers à hautes tiges</p>

			(fruits à pépins ou à noyau, noyers), des plantations de peupliers et de saules, et des aires afforestées dont la hauteur dominante est supérieure à 3 m, qui sont également mentionnés dans la description de “open forest”.
45-Brush forest	<b>Landcover --&gt; vegetated --&gt; Trees --&gt; Forest or Woodland</b>		<p>Cette catégorie englobe les zones de végétation naturelle semi-naturelle, avec des arbres comme forme de vie dominante. Cela correspond à la description de “brush forest” qui comprend des étendues boisées de manière homogène d’arbres et d’espèces arbustives de type buissonnant, à condition que le point d’échantillonnage soit à l’intérieur de la lisière du boisement, la hauteur dominante dépasse 3 m (exception: aussi moins de 3 m pour les aulnes verts et les pins rampants), le degré de couvert à l’intérieur de la surface de référence soit supérieur à 60%, la surface minimale soit supérieure à 25 m et que la proportion d’espèces non buissonnantes représente moins d’1/3.</p> <p>De plus, cette catégorie pourrait également inclure des espèces arbustives de type buissonnant comme l’aulne vert (<i>Alnus viridis</i>), le pin rampant (<i>Pinus mugo f. prostrata</i>), le noisetier commun (<i>Corylus avellana L.</i>), le genévrier commun (<i>Juniperus communis</i>), les saules de type rampant (<i>Salix sp.</i>), ainsi que des groupes d’arbres sur surfaces agricoles en plaine et en montagne, des groupes</p>

		d'arbres sur des surfaces improductives, des peuplements d'espèces forestières et d'espèces communes aux parcs (protections contre le vent et allées), des arbres fruitiers à hautes tiges (fruits à pépins ou à noyau, noyers) sur des surfaces d'habitat et d'infrastructure, des coins de forêt se terminant par des groupements d'arbres clairsemés, des aires afforestées dont la hauteur dominante est supérieure à 3 m et le degré de couvert inférieur à 20%, et des groupes d'arbres sur des coupes de bois et des surfaces forestières dévastées, qui sont également mentionnés dans la description de "brush forest".
46 - Linear woods	<b>Landcover --&gt; vegetated --&gt; Trees --&gt; Forest or Woodland</b>	Cette catégorie englobe les zones de végétation naturelle semi-naturelle, avec des arbres comme forme de vie dominante. Cela correspond à la description de "linear woods" qui comprend des peuplements d'arbres et d'arbustes d'au moins 25 m de long et de 25 m de large au maximum, à condition que le point d'échantillonnage soit situé sur le houppier, la hauteur dominante dépasse 3 m, et que le degré de couvert à l'intérieur de la surface de référence soit supérieur à 60%. De plus, cette catégorie pourrait également inclure des peuplements d'espèces forestières et d'espèces communes aux parcs (protections contre le vent et allées), des arbres fruitiers à

		hautes tiges (fruits à pépins ou à noyau, noyers) sur des surfaces d'habitat et d'infrastructure, des plantations de peupliers et de saules, des coins de forêt se prolongeant sur plus de 25 m de long, des peuplements d'arbres linéaires d'une hauteur supérieure à 3 m à proximité des aires afforestées, des coupes de bois ou des surfaces dévastées, et des haies taillées jusqu'à la souche sur les terres agricoles et sur les surfaces d'habitat et d'infrastructure, qui sont également mentionnés dans la description de "linear woods".
47 - Clusters of trees	<b>Landcover --&gt; vegetated --&gt; Trees --&gt; Forest or Woodland</b>	Cette catégorie englobe les zones de végétation naturelle semi-naturelle, avec des arbres comme forme de vie dominante. Cela correspond à la description de "clusters of tree" qui comprend des peuplements d'arbres comptant au minimum 3 arbres de plus de 3 m à l'intérieur de la lisière du boisement, ne remplissant pas les critères des catégories 41-46, mais dont le degré de couvert est inférieur à 20% (surface de référence) si la largeur est supérieure à 50 m, est inférieur à 60% (surface de référence) si la largeur est comprise entre 25 et 50 m, et peut prendre n'importe quelle valeur si la largeur est inférieure à 25 m. De plus, cette catégorie pourrait également inclure des groupes d'arbres sur surfaces agricoles en plaine et en montagne, des groupes d'arbres sur des surfaces improductives, des peuplements d'espèces

			forestières et d'espèces communes aux parcs (protections contre le vent et allées), des arbres fruitiers à hautes tiges (fruits à pépins ou à noyau, noyers) sur des surfaces d'habitat et d'infrastructure, des coins de forêt se terminant par des groupements d'arbres clairsemés, des aires afforestées dont la hauteur dominante est supérieure à 3 m et le degré de couvert inférieur à 20%, et des groupes d'arbres sur des coupes de bois et des surfaces forestières dévastées, qui sont également mentionnés dans la description de "clusters of tree".
50 - Bare land	51 - Solid rock	<b>Landcover --&gt; not vegetated --&gt; Natural --&gt; Natural Bare Surfaces --&gt; type of natural bare surface: Bare rocks</b>	La description de "Bare rocks" mentionne que la surface de la roche est continue, sauf peut-être pour quelques fissures dans le matériau. Certaines zones peuvent être recouvertes de couches superficielles de sol ou il pourrait y avoir des poches isolées de sol ou un mélange des deux. Cela semble correspondre à la description de "solid rock", qui inclut des surfaces rocheuses homogènes, des pans de rochers, des chantiers excavés jusqu'à la roche affleurante, etc.
	52 - Granular soil	<b>Landcover --&gt; not vegetated --&gt; Natural --&gt; Natural Bare Surfaces --&gt; type of natural bare surface: Gravel stones and/or boulders</b>	La description de "Gravel stones and/or boulders" mentionne que au moins 60% du total des fragments grossiers sont constitués de l'un des trois. Cela semble correspondre à la description de "granular soil", qui inclut des surfaces couvertes d'éboulis, de sable ou d'humus, des pierriers sous les

			parois rocheuses, des chantiers excavés jusqu'à la couche pierreuse, à la couche de sable ou d'humus, etc.
	53 - Rocky areas	<b>Landcover --&gt; not vegetated --&gt; Natural --&gt; Natural Bare Surfaces --&gt; type of natural bare surface: Gravel stones and/or boulders</b>	La description de "Gravel stones and/or boulders" mentionne que au moins 60% du total des fragments grossiers sont constitués de l'un des trois. Cela semble correspondre à la description de "rocky areas", qui inclut des surfaces couvertes d'une végétation herbacée ou d'une végétation buissonnante et parsemées de rochers saillants ou d'éboulis, de sable et d'humus, etc. Cependant, comme la description de "rocky areas"
60 - Water arcas	61 - Water	<b>Landcover --&gt; not vegetated --&gt; Natural --&gt; water, ice or snow --&gt; water</b>	
	62 - Glacier, perpetual snow	<b>Landcover --&gt; not vegetated --&gt; Natural --&gt; water, ice or snow --&gt; water/snow/ice</b>	
	63 - Wetlands	<b>Landcover --&gt; vegetated --&gt; no trees --&gt; not cultivated --&gt; life form: choisir selon la situation</b>	
	64 - Reedy marshes	<b>Landcover --&gt; vegetated --&gt; no trees --&gt; not cultivated --&gt; Grasses, rushes and/or sedges</b>	

## Script :

```
import sys
import json
import pathlib
import geopandas as gpd
import re

_CLASS_NAME_RE = re.compile("_(?P<name>.* )")

def _cleanup_classname(classname: str) -> str:
    m = re.search(_CLASS_NAME_RE, classname)
    extracted_classname = m.groupdict()["name"]
    return extracted_classname.strip().replace("_", " ").lower()

if __name__ == "__main__":
    file_path = pathlib.Path(sys.argv[1])

    with open(str(file_path)) as f:
        loaded_json = json.load(f)

    radiiuses = [ftr["radius"] for ftr in loaded_json["features"]]
    gdf = gpd.read_file(str(file_path))

    file_out = str(file_path.parent / f"{file_path.stem}_with_radius.gpkg")

    gdf["radius"] = radiiuses
    gdf["classname"] = gdf["comments"].apply(_cleanup_classname)

    gdf.to_file(file_out, driver="GPKG")
    print(f"File saved at {file_out}")
```

## Bibliographie :

Aberystwyth University. (n.d.). *EarthTrack - Download data*.

<https://earthtrack.aber.ac.uk/download/>

Chaturvedi, V., & de Vries, W. T. (2021). Machine learning algorithms for urban land use planning: A review. *Urban Science*, 5(3), 68.

<https://doi.org/10.3390/urbansci5030068>

Congedo, L. (2024). *Semi-Automatic Classification Plugin Documentation* (Version 8.3.0.2). <https://semiautomaticclassificationmanual.readthedocs.io/en/latest/>

Copernicus. (n.d.). Accueil. <https://www.copernicus.eu/fr>

Hermosilla, T., Ruiz, L. A., Recio, J. A., & Balsa-Barreiro, J. (2012). *Land-use mapping of Valencia city area from aerial images and LiDAR data*. In *Proceedings of the Fourth International Conference on Advanced Geographic Information Systems, Applications and Services (GEOProcessing 2012)* (pp. 92–97).

[https://www.researchgate.net/publication/236325319\\_Land-use\\_mapping\\_of\\_Valencia\\_city\\_area\\_from\\_aerial\\_images\\_and\\_LiDAR\\_data](https://www.researchgate.net/publication/236325319_Land-use_mapping_of_Valencia_city_area_from_aerial_images_and_LiDAR_data)

Office fédéral de la statistique. (2019). *La couverture du sol - Nomenclature de la couverture du sol NOLC04*. <https://www.bfs.admin.ch/asset/fr/11087257>

Sinergise. (2024). CORINE Land Cover — Sentinel Hub Collections.

<https://collections.sentinel-hub.com/corine-land-cover/readme.html>

SITG – Système d'information du territoire à Genève. (n.d.). *Carte interactive*.

<https://map.sitg.ge.ch/app/?mapresources>

TERN. (2023). *EarthTrack report – Final draft*. Living Earth Hub. <https://livingearthhub.org>

United States Geological Survey. (n.d.). *EarthExplorer*. <https://earthexplorer.usgs.gov/>

Université de Genève. (2024). *Certificat complémentaire en géomatique* [Cours en ligne]. Moodle UNIGE. <https://moodle.unige.ch/course/view.php?id=16488>