

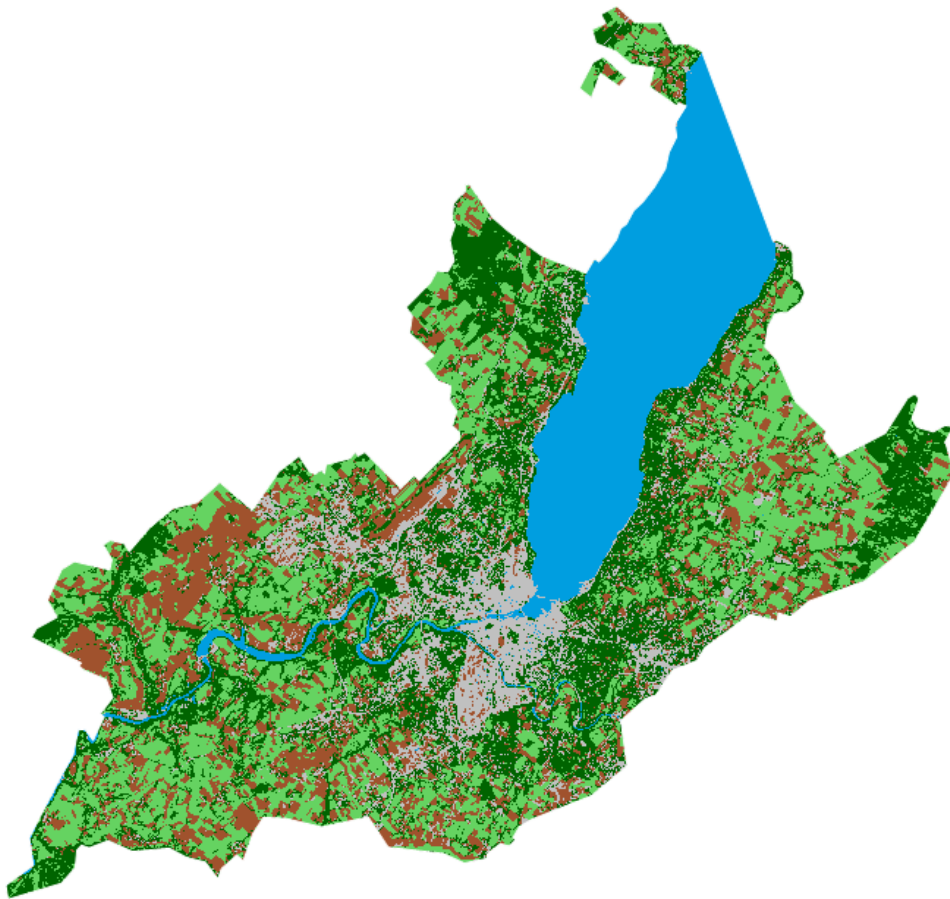


ÉCOLE POLYTECHNIQUE  
FÉDÉRALE DE LAUSANNE

ESPACES VERTS ET SANTÉ EN MILIEU URBAIN : ANALYSE DE LA  
RELATION ENTRE CLASSES DE COUVERTURE DU SOL ET L'INDICE  
DE MASSE CORPORELLE D'UN ÉCHANTILLON REPRÉSENTATIF DE LA  
POPULATION DU CANTON DE GENÈVE

PROJET SIE

LASIG



ESTELLE ROCHAT

JANVIER 2014

ENCADRANT :  
STÉPHANE JOOST



# Table des matières

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Données</b>	<b>2</b>
2.1	Données d'imagerie . . . . .	2
2.2	Statistique Suisse de la superficie . . . . .	4
2.3	Classification LiDAR de Florent Cornuet . . . . .	4
2.4	Données de santé . . . . .	5
<b>3</b>	<b>Méthode</b>	<b>6</b>
3.1	Délimitation de la zone d'étude . . . . .	7
3.2	Classification de la couverture du sol . . . . .	7
3.2.1	Choix des classes . . . . .	7
3.2.2	Classification avec Erdas Imagine . . . . .	8
3.2.3	Classification avec Matlab . . . . .	9
3.2.4	Calcul de la précision des résultats . . . . .	10
3.3	Comparaison avec la statistique Suisse de la superficie . . . . .	11
3.4	Comparaison avec la classification de Florent Cornuet . . . . .	11
3.5	Création d'une table avec les données de couverture du sol et les données de santé . . . . .	12
3.6	Analyse des interactions avec la santé . . . . .	13
<b>4</b>	<b>Résultats des classifications et discussion</b>	<b>14</b>
4.1	Classification avec Erdas Imagine . . . . .	14
4.2	Classification avec Matlab . . . . .	17
4.3	Précisions obtenues . . . . .	18
4.4	Comparaison avec la statistique de la superficie . . . . .	19
4.5	Comparaison avec la classification de Florent Cornuet . . . . .	20
4.6	Conclusion sur les classifications . . . . .	22
<b>5</b>	<b>Résultats de l'étude des interactions avec la santé et discussion</b>	<b>23</b>
5.1	Corrélations linéaires . . . . .	23
5.2	Régressions linéaires . . . . .	25
5.3	LISA bivariés . . . . .	26
5.4	Conclusion sur les interactions avec la santé . . . . .	29
<b>6</b>	<b>Conclusion</b>	<b>30</b>
<b>A</b>	<b>Annexe : Catégories de la statistique Suisse de la superficie</b>	<b>VIII</b>
<b>B</b>	<b>Annexe : Codes Matlab</b>	<b>IX</b>
B.1	Code Matlabs utilisé pour la classification . . . . .	IX
B.2	Code Matlab utilisé pour les calculs des précisions . . . . .	X
B.3	Code Matlab utilisé pour le calcul des pourcentages dans les cellules des grilles . . . . .	XI
<b>C</b>	<b>Annexe</b>	<b>XIII</b>
C.1	Correspondance entre les classes de la classification et les catégories de la statistique Suisse de la superficie . . . . .	XIII
C.2	Correspondance entre les classes de la classification de Florent Cornuet et les catégories de la statistique Suisse de la superficie . . . . .	XIV
<b>D</b>	<b>Annexe : Résultat des classifications</b>	<b>XV</b>
D.1	Annexe : Résultat de la classification avec Erdas . . . . .	XV
D.2	Annexe : Résultat de la classification avec Matlab . . . . .	XVI

<b>E</b>	<b>Annexe : Valeurs numériques des corrélations</b>	<b>XVII</b>
E.1	Adultes, Classification réalisée avec Matlab . . . . .	XVII
E.2	Adultes, Classification réalisée par Florent Cornuet . . . . .	XVIII
E.3	Enfants, Classification réalisée avec Matlab . . . . .	XIX
E.4	Enfants, Classification réalisée par Florent Cornuet . . . . .	XX



## Table des figures

2.1	Image SPOT multispectrale à 10 m de résolution utilisée pour la classification (fausses couleurs infrarouges).	3
2.2	Image SPOT panchromatique à 5m de résolution.	3
2.3	Exemple de grille d'échantillonnage superposée à une image aérienne. <a href="http://www.bfs.admin.ch/">http://www.bfs.admin.ch/</a>	4
2.4	Surface couverte par la classification de Florent Cornuet.	5
3.1	Démarche générale et logiciels utilisés	6
3.2	Zone d'étude	7
3.3	Exemple d'éléments de chaque classe sur l'image infrarouge (fausses couleurs)	8
3.4	Matrice de confusion	10
3.5	Décalage entre les pixels des deux classifications.	11
4.1	Réflectance moyenne des différentes classes sur les quatre bandes pour les échantillons donnés comme référence (réflectances données par Erdas).	14
4.2	Résultat de la classification réalisée avec Erdas Imagine (résolution : 10m).	14
4.3	Petits espaces verts correctement classifiés à l'intérieur du centre urbain dense.	15
4.4	Passage visible d'une zone urbaine dense à une zone plus verte.	15
4.5	Bordure des cours d'eau classée comme "urbain".	15
4.6	Rivière classée comme "urbain".	16
4.7	Classe "sol" apparaissant abondamment dans les zones urbaines de faible densité	16
4.8	Champs très lumineux classés comme zones urbaines.	17
4.9	Résultat de la classification réalisée avec Matlab.	17
4.10	Polygones manuellement classés utilisés pour estimer la précision.	18
4.11	Précisions des résultats obtenus avec Erdas et Matlab	18
4.12	Précisions obtenues en comparant les résultats des classifications avec la statistique Suisse de la superficie	19
4.13	Précisions obtenues en comparant les résultats des classifications avec la classification réalisée par Florent Cornuet	20
4.14	Différence du niveau de détail entre les deux classifications	21
4.15	Pixels classés comme eau dans les zones d'ombre.	21
4.16	Précisions obtenues en comparant la classification de Florent Cornuet avec les polygones digitalisés dans Erdas et avec la statistique Suisse de la superficie	21
5.1	Corrélations linéaires entre les données de couverture du sol obtenues avec la classification réalisée avec Matlab, et les données IMC des adultes (haut) et des enfants (bas)	23
5.2	Corrélations linéaires entre les données de couverture du sol obtenues avec la classification réalisée par Florent Cornuet, et les données IMC des adultes (haut) et des enfants (bas)	24
5.3	Régressions linéaires entre l'IMC corrigé et les pourcentages représentés par les classes de couverture du sol dans les cellules de 800m	25
5.4	Régressions linéaires entre l'IMC corrigé et les pourcentages représentés par les classes de couverture du sol dans les cellules de 800m, avec les données de la classification de Florent Cornuet	26
5.5	LISA bivariés entre le BMI corrigé des adultes et le pourcentage cumulé de végétation et forêt dans les cellules de 3200m, avec la classification réalisée avec Matlab.	26
5.6	LISA bivariés entre le BMI corrigé des adultes et le pourcentage cumulé de végétation, sol et forêt dans les cellules de 800m, avec la classification réalisée avec Matlab.	27
5.7	LISA bivariés entre le BMI corrigé des enfants et le pourcentage cumulé de surfaces de végétation, de sol et de forêt dans les cellules de 200m, avec la classification réalisée avec Matlab.	27

5.8	LISA bivariés entre le BMI corrigé des adultes et le pourcentage de surfaces perméables dans les cellules de 800m, avec la classification réalisée par Florent Cornuet.	28
5.9	LISA bivariés entre le BMI corrigé des enfants et le pourcentage cumulé de surfaces perméables et de végétation dans les cellules de 800m, avec la classification réalisée par Florent Cornuet. . . . .	28
5.10	LISA bivariés entre le BMI corrigé des enfants et le pourcentage de surfaces perméables dans les cellules de 200m, avec la classification réalisée par Florent Cornuet.	28

# 1 Introduction

Durant les dernières décennies, la croissance démographique s'est accompagnée d'une densification des villes, parfois au détriment des espaces verts situés à l'intérieur des zones urbaines. Pourtant, de nombreuses études ont montré que l'interaction avec la nature pouvait avoir un effet bénéfique sur la santé physique et mentale. En effet, les espaces verts situés à proximité des habitats peuvent encourager à passer plus de temps à l'extérieur et/ou augmenter l'activité physique [1], notamment en favorisant la marche et le vélo pour les loisirs et les déplacements. L'augmentation de l'activité physique liée aux espaces verts peut également être visible par son impact sur le poids des individus, le pourcentage de zones vertes diminuant ainsi les problèmes de surpoids [2, 3, 4, 5].

De plus, des études ont montré que l'exposition à l'environnement naturel pouvait avoir des effets positifs également sur la santé psychique, spirituelle et sociale. Les individus vivants dans des zones densément urbanisées auraient ainsi tendance à être plus isolés [6] et à présenter un risque plus élevé d'être sujet à des troubles mentaux [1]. A l'inverse l'interaction avec la nature favoriserait l'estime de soi [6], réduirait le stress [6, 3], favoriserait la concentration [1, 6], faciliterait les interactions sociales [6] et réduirait la violence [6]. Même la contemplation passive de la nature ou les interactions indirectes (présence de plantes vertes, tableaux de scènes naturelles...) pourraient avoir un effet bénéfique [1, 6].

A l'inverse d'autres études sont arrivées à la conclusion qu'aucune relation significative ne pouvait être mise en évidence entre l'espace naturel et la santé et par exemple que le poids des enfants ne pouvait pas être lié à l'environnement naturel [7]. Certaines recherches ont également conclu que ces interactions variaient selon l'âge des individus [1, 8, 4] ou qu'elles dépendaient de leur statut socio-économique [1, 9, 4]. De plus, des relations différences ont souvent été remarquées entre les milieux urbains denses et les milieux plus ruraux [8, 4].

Les interactions entre la santé et l'espace naturel sont donc complexes et les méthodes utilisées pour les mettre en évidence sont très variées. Les données liées aux espaces verts peuvent fortement varier, allant de la présence ou absence d'un jardin privé [1] à une classification complète de la couverture du sol [1, 8, 9]. De plus, différentes distances peuvent être prises en compte lors de l'étude du voisinage. Les données de santé utilisées peuvent également être très différentes, comprenant des sondages et questionnaires liés à la santé [1, 8, 9], des études sur le nombre de jours passés à l'hôpital [6], ou encore des statistiques liées aux causes de mortalité [6]. Finalement, l'interaction entre les deux types de données peut être étudiée avec différentes méthodes, allant de l'utilisation de régressions linéaires multiples calculées entre les indicateurs de santé et les paramètres liés aux espaces verts [1, 9] à la répétition successive d'expériences en présence ou absence de végétation [6].

Dans ce contexte, l'objectif de ce projet est d'analyser l'interaction entre la couverture du sol et la santé, plus précisément l'indice de masse corporelle, dans le canton de Genève, afin de voir si des relations similaires à celles mentionnées plus haut peuvent être mises en évidence. Le premier objectif est donc de produire des données de couverture du sol en réalisant une classification à partir d'images aériennes ou satellitaires. Les résultats devront permettre de calculer le pourcentage représenté par chaque classe de couverture du sol dans des cellules de 200m, 800m et 3200m de côté. Le but est également de comparer les résultats de cette classification avec les données fournies par la statistique Suisse de la superficie et avec une classification réalisée par Florent Cornuet [10] à partir de données LiDAR.

Le second objectif est d'étudier la relation entre ces données et l'indice de masse corporelle d'adultes et d'enfants domiciliés dans le canton de Genève. L'analyse sera réalisée successivement avec les grilles de 200m, 800m et 3200m de côté, en utilisant des calculs de corrélations linéaires, des régressions linéaires et des graphiques LISA bivariés.

## 2 Données

Différentes données ont été utilisées dans ce projet et peuvent être regroupées en quatre sections : les données d'imagerie nécessaires pour la classification de la couverture du sol, les données de la statistique Suisse de la superficie, le résultat d'une classification réalisée par Florent Cornuet et les données de santé.

### 2.1 Données d'imagerie

Différentes images aériennes et satellitaires étaient disponibles au LASIG pour permettre la classification de la couverture du sol dans le canton de Genève :

- Les orthophotos swissimage panchromatiques à 50cm de résolution de 1998, 2004 et 2006.
- Les orthophotos swissimage panchromatiques à 25cm de résolution de 2006 et 2009.
- Les orthophotos swissimage infrarouges à 25cm de résolution de 2012.
- Les images satellitaires multispectrales LANDSAT à 30m de résolution.
- Les images satellitaires panchromatiques SPOT5 à 5m de résolution de 2004.
- Les images satellitaires multispectrales SPOT5 à 10m de résolution de 2004.

Le choix s'est porté sur les données satellitaires SPOT qui possédaient la résolution spatiale la plus adaptée pour ce projet. En effet, les orthophotos ayant une résolution très grandes, les fichiers auraient été lourds à traiter. A l'inverse la résolution des images LANDSAT était trop faible pour assurer une classification suffisamment précise, prenant en compte les petits espaces verts au centre des milieux urbains. Les images SPOT constituaient donc une alternative intéressante en présentant une résolution spatiale moyenne bien adaptée aux besoins et une bande infrarouge nécessaire pour la classification de la végétation.

Les satellites SPOT (Satellite Pour l'Observation de la Terre) sont des satellites de télédétection développés par l'agence spatiale française (CNES - Centre national d'études spatiales) [11]. Les images fournies par ces satellites sont commercialisées par la compagnie Astrium, rattachée à l'EADS (European Aeronotic Defence and Space Company) [12]. Le satellite SPOT5, depuis lequel ont été prises les images à disposition pour ce projet, a été mis en fonction en 2002. Ce satellite permet d'acquérir des images panchromatiques à 5 m de résolution et des images multispectrales à 10 m de résolution, possédant les 4 bandes suivantes [13] :

- Bande 1 : Vert (G) : 0.5 - 0.59  $\mu\text{m}$
- Bande 2 : Rouge (R) : 0.61 - 0.68  $\mu\text{m}$
- Bande 3 : Proche infrarouge (NIR) : 0.78 - 0.89  $\mu\text{m}$
- Bande 4 : Moyen infrarouge (MIR) : 1.58 - 1.75  $\mu\text{m}$

Tout le canton de Genève est couvert par une image multispectrale et deux images panchromatiques datant du 25 mai 2004 [14] . La classification a été réalisée sur la base de l'image multispectrale, déjà corrigée des effets atmosphériques et des différences d'illumination [13]. Les caractéristiques de cette image sont résumées dans le tableau ci-dessous.

Format	PIX
Projection	CH1903+
Date de la prise de vue	29.05.2004
Résolution spatiale	10 x 10m
Résolution spectrale	4 bandes
Résolution radiométrique	8 bit

Zone couverte	Xmin : 478'625	Ymin: 142'325
	Xmax: 519'335	Ymax: 96'985
Largeur	40'710 m	4071 pixels
Hauteur	45'340 m	4534 pixels
Surface	1845.8 km <sup>2</sup>	18'457'914 pixels

Cette image contenait des bandes supplémentaires avec des données n'étant pas des réflectances (azimut solaire, zénith, visibilité,...). Ces bandes ont donc été supprimées et seules les bandes 1 à 4, contenant les réflectances, ont été conservées pour la suite du travail. La figure 2.1 présente l'image utilisée pour la classification (pour la zone d'étude retenue, voir section 3.1)

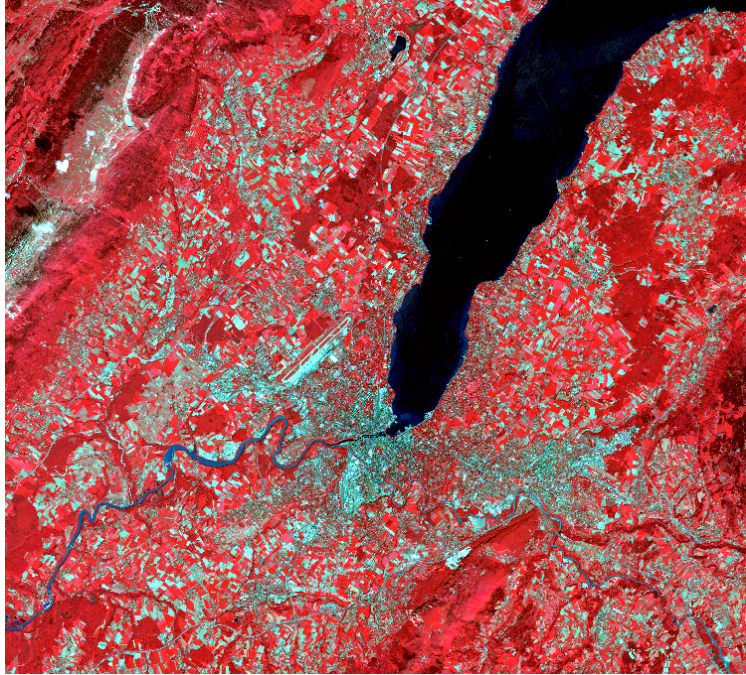


FIGURE 2.1 – Image SPOT multispectrale à 10 m de résolution utilisée pour la classification (fausses couleurs infrarouges).

Les images panchromatiques à 5m de résolution (figure 2.2) ont été utilisées comme aide lors du dessin des polygones servant de base à la classification supervisée et lors de la création des polygones utilisés pour estimer la précision du résultat de la classification (voir chapitre 3 Méthode).



FIGURE 2.2 – Image SPOT panchromatique à 5m de résolution.



## 2.2 Statistique Suisse de la superficie

Les données de la statistique Suisse de la superficie ont été utilisées pour une comparaison avec les résultats de la classification. Ces données sont créées par l'Office Fédéral de la Statistique (OFS) et constituent un relevé de la couverture du sol par échantillonnage à partir d'images aériennes. Les points d'échantillonnages sont définis par des angles droits sur une grille régulière hectométrique (voir figure 2.3). Cette grille est superposée aux images aériennes et chaque point est observé par un opérateur qui attribue une classe au pixel situé sous le point d'échantillonnage. Pour les points où l'attribution est incertaine, une visite sur le terrain est ensuite réalisée afin de contrôler l'attribution. [15]



FIGURE 2.3 – Exemple de grille d'échantillonnage superposée à une image aérienne. <http://www.bfs.admin.ch/>

Il s'agit donc de données ponctuelles ne pouvant pas être utilisées pour une représentation sous forme de carte. En effet, le point d'échantillonnage peut tomber sur une petite surface représentée par un bâtiment isolé, une route ou un cours d'eau et il ne serait évidemment pas correct d'appliquer cette classe à tout l'hectare.

Les données utilisées datent de 2009 et comprennent 72 catégories de couverture du sol, hiérarchisées dans un système arborescent comprenant au départ quatre domaines principaux : surfaces boisées, surfaces agricoles, surfaces d'habitat et d'infrastructures et surfaces improductives. Un tableau détaillé de toutes les classes est présent dans l'Annexe A. Ces données étaient disponibles dans un fichier vectoriel comprenant les points d'échantillonnage et les catégories associées.

## 2.3 Classification LiDAR de Florent Cornuet

Une classification de la couverture du sol dans le canton de Genève a été réalisée par Florent Cornuet en 2011 dans le cadre de son projet de master (voir [10]). Cette classification a été réalisée sur la base de données LiDAR datant de 2009. Le principe du LiDAR est d'envoyer depuis un émetteur aéroporté une onde qui sera réfléchiée par la surface du sol. La mesure du temps de parcours aller et retour de l'onde et la connaissance de la position de l'émetteur permettent de déduire l'élévation du point au sol. A partir de ces données, des modèles numériques de terrain (MNT) et des modèles numériques de surface (MNS), contenant l'élévation de tous les points du sol, respectivement des points des éléments au-dessus du sol (bâtiment, arbres,...), peuvent être créés. La classification réalisée par Florent Cornuet s'est appuyée sur la combinaison d'un MNS, d'un MNT, du nuage brut de points LiDAR (contenant notamment des informations sur

l'intensité du rayonnement réfléchi et le type d'écho) et d'une carte 2D du bâti [10]. Des images numériques ont été créées sur la base de ces informations et des techniques de traitement d'image ont ensuite permis de réaliser la classification [10].

La résolution finale de la carte classifiée est de 1m, ce qui correspond à la résolution des données LiDAR. La classification comprend 5 classes [10] :

1. **Eau** : pixels ayant une hauteur (MNT) valant 0 et une intensité nulle.
2. **Sol perméable** : pixels ayant une hauteur inférieure à 1.5 m et une intensité supérieure à 30. Cette classe contient donc les champs, les prairies ou les surfaces contenant de la végétation basse.
3. **Bâtiment** : pixels ayant une hauteur (MNS) supérieure à 1m et correspondant aux zones définies dans la carte 2D du bâti.
4. **Végétation** : pixels ayant une hauteur supérieure à 1 m et n'étant pas classifiés comme bâtiments.
5. **Sol imperméable** : pixels ayant une hauteur inférieure à 1.5 m et une intensité inférieure à 30. Cette classe comprend les routes, l'asphalte et les surfaces bétonnées.

La zone couverte par cette classification est présentée sur la figure 2.4

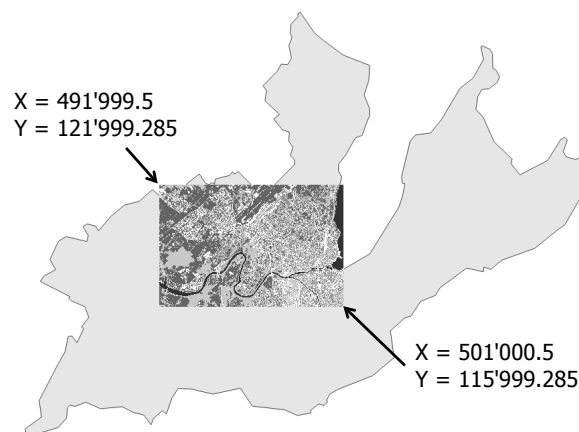


FIGURE 2.4 – Surface couverte par la classification de Florent Cornuet.

Cette classification était disponible sous forme d'image au format .tif. Elle a été utilisée pour une comparaison avec les résultats de la classification réalisée sur la base des images SPOT.

## 2.4 Données de santé

Les données de santé utilisées pour ce projet sont issues des données récoltées par les bus de la santé dans le canton de Genève. Les bus de la santé sont un projet des Hôpitaux Universitaires Genevois (HUG) permettant de récolter des données de santé sur un échantillon d'individus représentatif de la population du canton, afin de surveiller l'évolution de facteurs à risques et de mener des campagnes de prévention [16]. Les individus sont sélectionnés aléatoirement mais de façon à ce que les classes d'âges, les localisations des domiciles, les emplois, etc. soient représentatifs de l'ensemble de la population [16]. Les personnes sélectionnées reçoivent un questionnaire concernant leur mode de vie (histoire médicale, alimentation, activité physique, tabagisme...) et sont invitées à réaliser une visite médicale au bus de la santé [17]. Ces visites permettent de récolter des données médicales (taille, poids, tension artérielle, fréquence cardiaque, santé mentale, prise de sang,...) permettant de mettre en évidence des facteurs à risques, liés notamment aux maladies cardiovasculaires [16],[17].

Les données à disposition pour ce projet regroupent les informations concernant 6663 adultes âgés de 34 à 74 ans et 3601 enfants âgés de 4 à 8 ans, tous domiciliés dans le canton de Genève. Elles comprennent notamment l'âge de chaque individu, l'adresse et les coordonnées du domicile, des informations liées à la situation socio-économique et l'indice de masse corporelle (IMC ou BMI). Cet indice correspond au rapport du poids (en kg) sur le carré de la taille (en m). Il constitue un indicateur permettant de dépister la surcharge pondérale qui est définie à partir d'un seuil de 30 kg/m<sup>2</sup>.

Des études précédentes ont montré que sur le jeu de données utilisé, une corrélation non négligeable pouvait être observée entre le revenu et l'indice de masse corporelle. De plus, des recherches ont déjà montré que la relation entre la santé et l'environnement naturel pouvait varier selon la situation socio-économique des individus considérés [1, 9, 4]. Par conséquent, afin de pouvoir mieux distinguer l'impact de l'environnement naturel, des données contenant l'IMC corrigé en fonction du revenu médian ont également été utilisées. Ces différents jeux de données étaient disponibles dans des fichiers vectoriels contenant un point pour chaque individu avec les adultes et les enfants sur des fichiers séparés.

### 3 Méthode

La figure 3.1 présente les grandes étapes réalisées dans ce projet et les logiciels utilisés. Chaque étape fait l'objet d'un sous-chapitre.

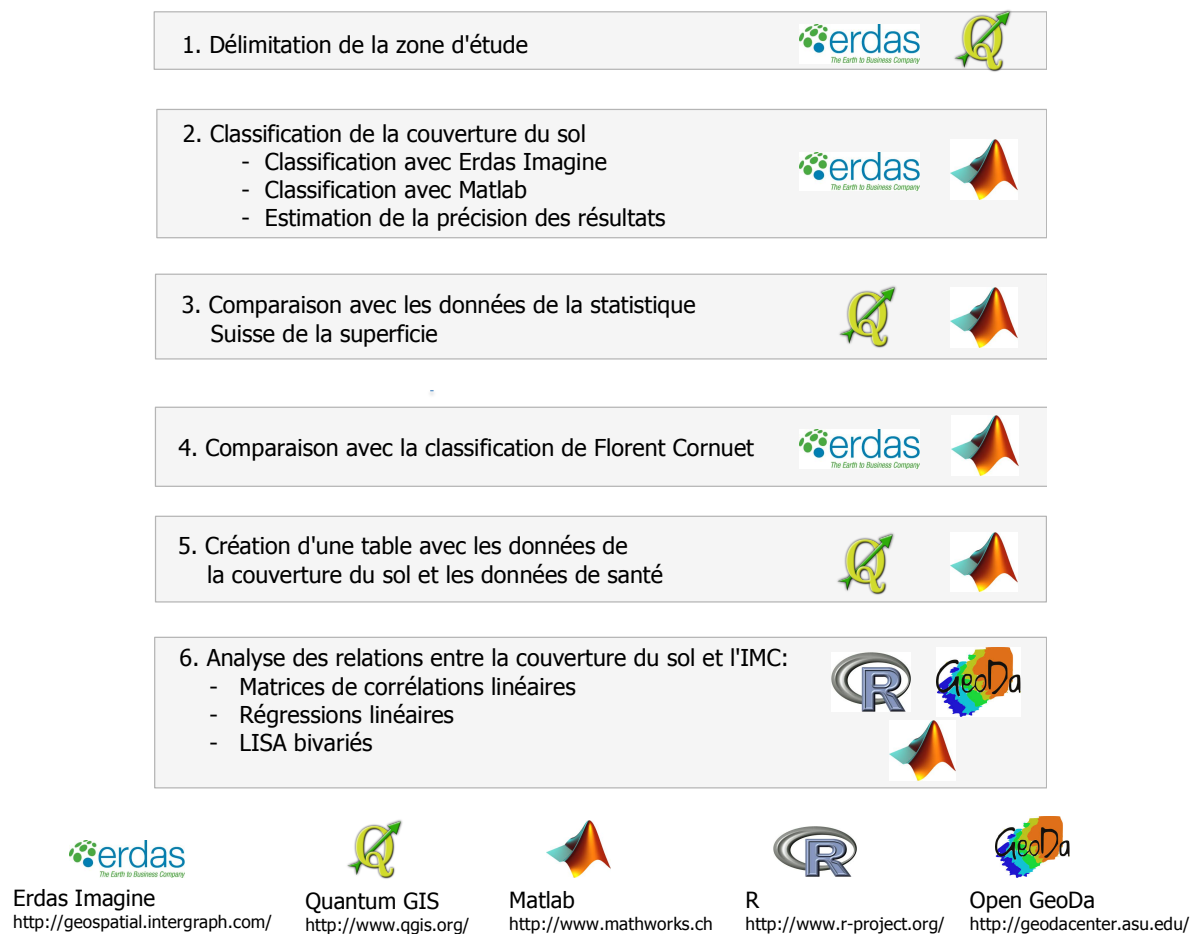


FIGURE 3.1 – Démarche générale et logiciels utilisés



### 3.1 Délimitation de la zone d'étude

La zone d'étude a été définie afin de couvrir l'entier du canton de Genève et de correspondre à un nombre entier de cellules de 3200 m. La zone retenue couvre une surface de 921.6 km<sup>2</sup> (9'216'000 pixels) présentée sur la figure 3.2.

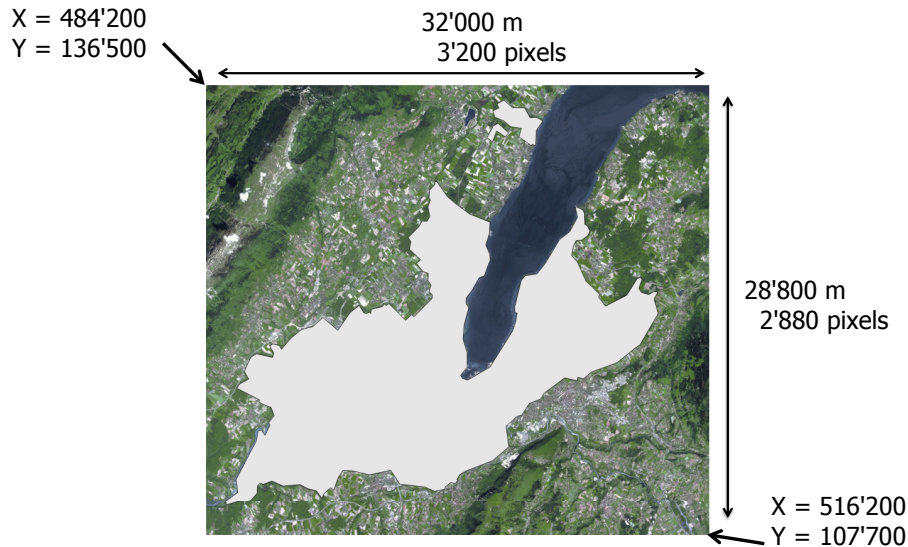


FIGURE 3.2 – Zone d'étude

Les grilles de 200m, 800m et 3200m ont été définies afin de couvrir exactement cette surface (Quantum GIS) et les données d'imagerie (images SPOT) ont été rognées afin de correspondre à cette région (Erdas). Les images classifiées correspondent également cette zone d'étude.

### 3.2 Classification de la couverture du sol

#### 3.2.1 Choix des classes

Le choix des classes a été fait de façon à ce que les réflectances des bandes spectrales à disposition permettent de discriminer chaque classe. La couverture du sol a ainsi été divisée en cinq classes distinctes :

1. **Végétation basse** : essentiellement les surfaces agricoles couvertes de végétation, mais aussi les terrains de foot, pelouses, parcs, etc. Tous ces éléments ont été regroupés dans la même classe car il est difficile de les différencier sur la base de leur réflectance.
2. **Eau** : toutes les surfaces contenant de l'eau (lacs, cours d'eau, marais, piscines, ...).
3. **Forêt** : surfaces forestières et arbres isolés.
4. **Sol nu** : généralement les sols agricoles non couverts par la végétation. Bien que l'utilisation du sol soit ici liée à l'agriculture au même titre que les surfaces végétalisées de la classe 1, ces surfaces ont été classées dans une catégorie différente car l'absence de végétation rend leur réflectance très différente. Cette classe contient également les autres surfaces perméables (chemins en terre battue, surfaces en graviers, ...).
5. **Urbain** : bâtiments et zones imperméables (routes, voies ferrées, ...). Différents tests ont été réalisés afin de différencier les bâtiments des autres zones imperméables, mais la résolution spatiale et les faibles différences de réflectance n'ont pas permis d'obtenir des résultats satisfaisants. Ces différents objets ont donc été maintenus dans la même classe, ce qui ne devrait pas avoir beaucoup d'effets sur l'analyse de l'interaction avec la santé.

La figure 3.3 donne des exemples de régions couvertes par les différentes classes sur l'image infrarouge affichée en fausses couleurs (Rouge = NIR, Vert = Rouge, Rouge = Bleu).

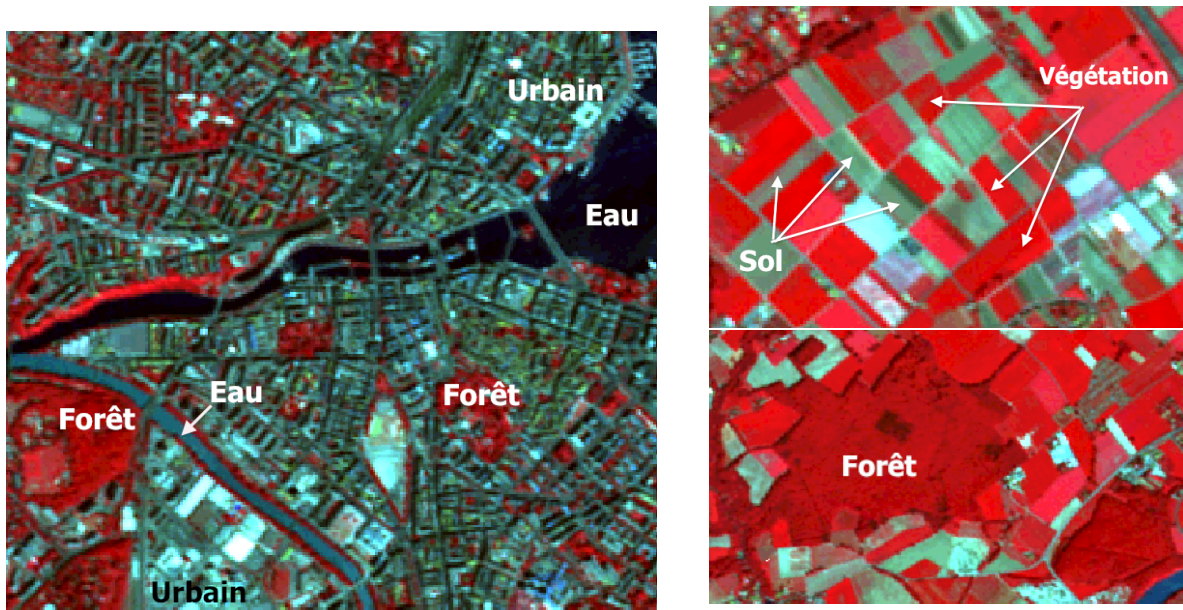


FIGURE 3.3 – Exemple d’éléments de chaque classe sur l’image infrarouge (fausses couleurs)

### 3.2.2 Classification avec Erdas Imagine

Erdas Imagine est un logiciel de télédétection développé par la compagnie Intergraph [18]. Ce logiciel permet un grand nombre d’opérations sur des images raster, notamment l’ajustement des contrastes et de la luminosité, l’application de filtres ou l’extraction d’indices particuliers. Il permet également de réaliser des classifications supervisées et non-supervisées à partir d’images aériennes ou satellitaires.

Dans le cadre de ce projet, Erdas a été utilisé afin de réaliser des classifications supervisées de la couverture du sol. D’autres logiciels tels qu’eCognition permettent également de réaliser un travail similaire. La différence principale entre les deux logiciels réside dans le fait qu’Erdas réalise la classification par pixel alors qu’eCognition segmente l’image en polygones de pixels similaires puis classe chaque polygone.

La classification réalisée est une classification supervisée qui s’appuie sur un échantillon de référence classé par l’utilisateur. A partir de cette référence, le logiciel classe les pixels selon leur ressemblance avec chaque classe prédéfinie. La première étape est donc de dessiner pour chaque classe des polygones contenant un échantillon de pixels de cette classe (fichier .aoi dans Erdas). Ces polygones ont été choisis au hasard, mais de façon à être répartis sur l’ensemble du territoire, à l’intérieur du canton de Genève. Ils ont été digitalisés sur la base des images panchromatiques à 5m de résolution.

Différentes méthodes peuvent ensuite être utilisées pour attribuer une classe aux pixels restants. Le logiciel Erdas offre la possibilité d’utiliser des méthodes paramétriques (faisant une hypothèse sur la distribution de la classe) ou non-paramétriques. Parmi les méthodes paramétriques, différentes possibilités sont proposées :

- Distance minimale : Sur la base des réflectances des échantillons donnés en exemple, le centroïde de chaque classe est calculé. Les pixels à classer sont ensuite assignés à la classe pour laquelle la distance spectrale au centroïde (distance euclidienne) est la plus faible [19].
- Maximum de vraisemblance gaussienne : Méthode basée sur l’hypothèse que les classes ont une distribution gaussienne. Les pixels à classer sont alors assignés à la classe pour

laquelle ils ont la plus grande probabilité d'appartenance. La méthode par défaut considère que la probabilité a priori d'appartenance à une classe est égale pour toutes les classes. Il est ensuite possible de définir des probabilités a priori différentes pour chaque classe (par exemple, donner une probabilité plus faible aux classes qui sont très peu présentes sur la zone d'étude) [19].

- Distance de Manahalonis : Méthode similaire à la méthode de la distance minimale, mais utilisant la matrice de covariance. Pour chaque classe, la variance des pixels donnés en échantillon est calculée. Lors de la classification des pixels restants, la variance des classes est prise en compte de telle sorte que la variance finale de la classe reste proche de celle calculée avec les pixels de l'échantillon de départ (par exemple, la variance des pixels d'une classe urbaine peut être plus grande que celle d'une classe d'eau) [19].

Les différentes méthodes ont été testées et les résultats ont été comparés. Des essais ont également été réalisés en essayant d'ajouter des classes supplémentaires (par exemple en distinguant les surfaces imperméables), en changeant les probabilités a priori d'apparition de chaque classe et en ajoutant des polygones dans la carte de vérité du sol. Les classifications qui semblaient les plus satisfaisantes ont ensuite été converties en image .tif couvrant la zone d'étude et ont été importées dans Matlab afin d'estimer la précision du résultat (voir 3.2.4). La méthode qui a finalement permis d'obtenir les meilleurs résultats est la méthode du maximum de vraisemblance gaussienne, avec des probabilités a priori d'appartenance conservées équivalentes pour toutes les classes.

### 3.2.3 Classification avec Matlab

Une classification supervisée a également été réalisée avec Matlab afin de comparer les résultats obtenus. A nouveau, pour réaliser cette classification, un échantillon de pixels correctement classés était nécessaire. Ces polygones ont été digitalisés dans Erdas sur des couches vectorielles séparées pour chaque classe. Les fichiers vectoriels ont ensuite été convertis en fichier raster au format .tif afin de pouvoir être importés dans Matlab où une carte présentant une "vérité du sol" a ainsi été établie. Dans Matlab, une sélection aléatoire a permis de retenir 1/5 des pixels ainsi classifiés afin de les utiliser comme base pour la classification supervisée. Les autres pixels ont été conservés pour tester ultérieurement la précision de la classification (voir 3.2.4).

La classification a été réalisée avec la fonction *classify* de Matlab. Différentes méthodes peuvent être utilisées avec cette fonction et sont présentées dans [20]. A nouveau, les différentes méthodes ont été testées et la méthode qui permettait d'obtenir les meilleurs résultats a été retenue. Il s'agissait de la méthode *quadratic* qui est basée sur l'ajustement pour chaque classe d'une courbe gaussienne, avec des variances différentes.

Des essais ont également été réalisés en appliquant des filtres morphologiques (dilatation, érosion,...) sur l'image et en stockant les résultats comme des bandes supplémentaires de l'image. Ces bandes supplémentaires permettent parfois de mieux discriminer les classes et donc d'améliorer la précision de la classification. Ce n'était toutefois pas le cas ici et ces bandes n'ont donc pas été conservées.

Le résultat de la classification était stocké dans une matrice sous Matlab. Afin de pouvoir utiliser ce résultat dans un SIG, un fichier .tif a été créé. Pour ce faire, les résultats ont tout d'abord été exportés dans un fichier .csv possédant trois colonnes : les coordonnées x et y du centre du pixel et la classe assignée au pixel. Ce fichier .csv a ensuite été importé dans Quantum GIS et a permis de créer un fichier vectoriel de points avec un point situé au centre de chaque pixel. La fonction *rasterize* de Quantum GIS a finalement permis de transformer ce fichier vectoriel en fichier raster au format .tif.

Les codes Matlab utilisés pour les différentes étapes sont présentés dans l'Annexe B.1 .

### 3.2.4 Calcul de la précision des résultats

L'analyse de la précision des différentes classifications a été réalisée dans Matlab sur la base de la carte de "vérité du sol" créée en digitalisant des polygones sur Erdas. Pour estimer la précision de la classification d'Erdas, les polygones qui ont été utilisés sont les mêmes que ceux créés pour la classification avec Matlab (ces polygones étant différents de ceux digitalisés sur les fichiers .aoi pour la classification avec Erdas). Pour la classification de Matlab, ces mêmes polygones ont été utilisés mais uniquement avec les pixels n'ayant pas été retenus pour la classification.

A partir de ces échantillons manuellement classés, des matrices de confusion (CM) ont été calculées, permettant de mettre en évidence pour chaque classe le nombre de pixels correctement classés et le nombre de pixels mal classés (voir figure 3.4) .

Classe de la classification \ Classe selon les polygones de référence	Classe selon les polygones de référence				
	1	2	3	4	5
1	$n_{11}$	$n_{12}$	$n_{13}$	$n_{14}$	$n_{15}$
2	$n_{21}$	$n_{22}$	$n_{23}$	$n_{24}$	$n_{25}$
3	$n_{31}$	$n_{32}$	$n_{33}$	$n_{34}$	$n_{35}$
4	$n_{41}$	$n_{42}$	$n_{43}$	$n_{44}$	$n_{45}$
5	$n_{51}$	$n_{52}$	$n_{53}$	$n_{54}$	$n_{55}$

FIGURE 3.4 – Matrice de confusion

Divers indicateurs de précisions ont ensuite été calculés à partir de ces matrices :

— **Overall accuracy (OA)** : pourcentage total de pixels correctement classifiés.

$$OA = \frac{\sum n_{ii}}{N} \text{ où } N \text{ est le nombre total de pixels utilisés dans la CM}$$

— **User accuracy (UA)** : Pour une classe  $i$ , cet indicateur donne la probabilité qu'un pixel classé en  $i$  appartienne réellement à la classe  $i$  (càd : pourcentage de pixels correctement classés par rapport au nombre total de pixels classés dans cette classe).

$$\text{Pour la classe A : } UA_A = \frac{n_{AA}}{\sum n_{Ai}}$$

— **Producer accuracy (PA)** : Pour une classe  $i$ , probabilité qu'un pixel appartenant réellement à cette classe soit classé en  $i$  (càd : pourcentage de pixels correctement classés par rapport au nombre total de pixels appartenant réellement à cette classe).

$$\text{Pour la classe A : } PA_A = \frac{n_{AA}}{\sum n_{iA}}$$

— **Indice kappa ( $\kappa$ )** : Cet indice exprime le pourcentage de réduction entre l'erreur générée par la classification et l'erreur qui serait générée par une classification totalement aléatoire. Cet indice présente l'avantage d'être insensible au nombre de pixels présents dans chaque classe.

$$\kappa = \frac{p_{obs} - p_{est}}{1 - p_{est}} \text{ avec } p_{est} = \frac{\sum_i \frac{\sum_j n_{ij} \sum_j n_{ji}}{N}}{N} \text{ et } p_{obs} = \frac{\sum_i n_{ii}}{N} = OA$$

Le code Matlab utilisé pour ces calculs est présenté dans l'Annexe B.2.

### 3.3 Comparaison avec la statistique Suisse de la superficie

Les résultats des classifications ont été comparés avec la statistique de la superficie en extrayant pour chaque point de la statistique la classe correspondante sur la classification, grâce à l'outil *Point Sampling Tool* de Quantum GIS. Comme la classification comporte cinq classes, les 72 catégories de la statistique de la superficie ont du être agrégées dans ces cinq classes afin de permettre la comparaison. Toutefois, pour plusieurs catégories de la statistique de superficie, l'attribution d'une classe n'était pas évidente et ces points ont donc été ignorés pour la comparaison. La répartition des catégories selon les cinq classes est présentée dans l'Annexe C.1. Pour cette comparaison, les classes "sol" et "végétations basses" ont été regroupées car en raison de leur différence liée à la période de culture, elles ne sont pas différenciées dans la statistique de la superficie.

Un code Matlab a ensuite permis de calculer une matrice de confusion et les différents paramètres statistiques présentés dans la section 3.2.4 (OA, UA, PA,  $\kappa$ ). Ce code est similaire à celui de la partie précédente (voir Annexe B.2).

### 3.4 Comparaison avec la classification de Florent Cornuet

Les résultats des classifications ont également été comparés avec la classification réalisée par Florent Cornuet. La première étape pour permettre cette comparaison a été d'ajuster les zones couvertes par les deux classifications. En effet, les résolutions des deux images étaient différentes et la zone couverte par la classification de Florent Cornuet était plus petite. De plus, les intersections des pixels des deux classifications ne correspondaient pas (un décalage de 0.5m était présent). Cette différence est présentée sur la figure 3.5.

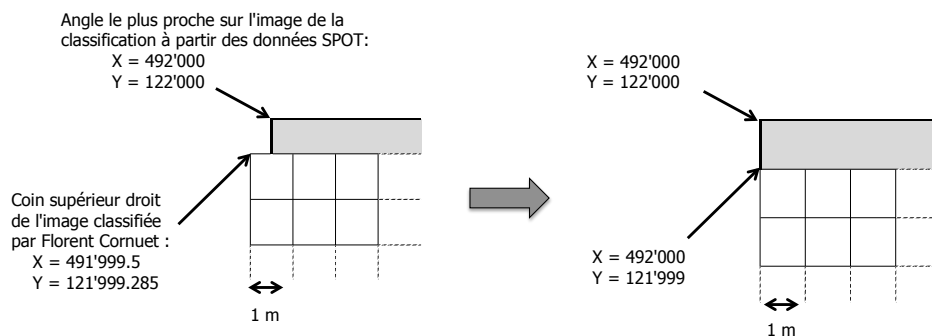


FIGURE 3.5 – Décalage entre les pixels des deux classifications.

L'image de Florent Cornuet a donc été importée dans Erdas et la fonction *Create Subset Image* a permis de recalculer par interpolation les valeurs des pixels déplacés de 0.5m (voir figure 3.5). L'image de la classification réalisée avec les données SPOT a ensuite été ré-échantillonnée pour obtenir des pixels de 1m, avec la fonction *Resample Pixel Size* d'Erdas (les 100 pixels qui étaient situés dans une cellule de 10x10m obtiennent la même valeur). Enfin, cette dernière image a été coupée pour couvrir exactement la même surface que celle de la classification de Florent Cornuet.

Afin de comparer les deux classifications, une correspondance entre les classes a également dû être établie. La classe "eau" était présente dans les deux classifications. La classe "urbain" réalisée sur la base des images SPOT regroupe les classes "bâtiments" et "surfaces imperméables" de la classification de Florent Cornuet. La classe "forêt" devrait correspondre à la classe "végétation" de Florent Cornuet puisque cette dernière a été définie comme la végétation de hauteur supérieure à 1.5m. Enfin, les classes "végétations basses" et "sol" sont regroupées dans la catégorie "sol perméable" chez Florent Cornuet. Le tableau ci-dessous résume les correspondances.

Classification réalisée à partir des images SPOT	Classification réalisée par Florent Cornuet
Eau	Eau
Urbain	Bâtiment Sol imperméable
Forêts	Végétation
Végétation basse Sol	Sol perméable

Finalement, un code Matlab similaire à ceux des deux sections précédentes (voir Annexe B.2) a permis de calculer les matrices de confusion et les paramètres statistiques présentés dans la section 3.2.4 (OA, UA, PA,  $\kappa$ ). Afin de comparer les résultats, les indices de précision ont également été calculés pour les comparaisons entre la classification réalisée par Florent Cornuet et les polygones manuellement digitalisés comme "vérité du sol", et entre la classification de Florent Cornuet et les données de la statistique Suisse de la superficie. Pour cette dernière comparaison, les correspondances utilisées entre les classes et les catégories de la statistique de la superficie sont présentées dans l'Annexe C.2.

### 3.5 Création d'une table avec les données de couverture du sol et les données de santé

L'analyse de l'interaction entre les données de santé et la couverture du sol a été réalisée avec la classification qui présentait les meilleurs résultats selon les calculs de précisions et les comparaisons avec la statistique de la superficie et la classification de Florent Cornuet. Les résultats de cette classification ont tout d'abord été agrégés dans les cellules des grilles de 200m, 800m et 3200m. Pour ce faire, un code Matlab a permis de calculer pour chaque cellule le pourcentage de surface représenté par chaque classe de couverture du sol (code disponible dans l'Annexe B.3). Pour chaque grille, les résultats ont été reportés dans un fichier .csv possédant 10 colonnes : les coordonnées x et y du centre de la cellule (2 colonnes), le pourcentage représenté par les 5 classes de couverture du sol (5 colonnes), le pourcentage cumulé des surfaces de végétation et de sol (1 colonne), de végétation et de forêt (1 colonne) et finalement de végétation, sol et forêt (1 colonne). Ces dernières agrégations ont été réalisées afin de mieux représenter les classes de végétation. En effet, la distinction entre la classe "sol" qui contient les champs non couverts par la végétation et la classe "végétation basse" dépend notamment de la date des prises de vue. Une catégorie regroupant ces deux classes peut donc permettre de mieux représenter le pourcentage de végétation basse, indépendamment de la saison. Les deux autres agrégations ont été réalisées afin de regrouper les classes d'éléments naturels et de voir si ces catégories agrégées permettaient une meilleure mise en évidence des interactions avec la santé.

Les fichiers .csv ont ensuite été importés dans Quantum GIS et convertis en fichiers vectoriels de points, comprenant un point au centre de chaque cellule de la grille. L'outil *Join by location* a permis de transférer les attributs de ces points vers les cellules (polygones) des grilles. Enfin, l'outil *Point Sampling Tool* a permis d'extraire pour chaque point où une donnée IMC était présente les informations de la cellule sur laquelle le point était situé. Cette démarche a été répétée pour toutes les grilles et pour les données d'IMC et d'IMC corrigé afin d'obtenir deux fichiers vectoriels finaux, l'un pour les adultes, l'autre pour les enfants, contenant les données de santé, l'IMC corrigé et les pourcentages de chaque classe pour les différentes grilles.

Finalement, deux tables ont été créées de façon similaire en n'utilisant non pas les résultats de la classification réalisée pour ce projet, mais en utilisant les résultats de la classification de Florent Cornuet. Toutefois, dans ces deux tables, les informations des cellules de 3200 m n'ont pas été reportées car seules deux cellules de cette grille étaient couvertes par la classification de Florent Cornuet. Cette étape a été réalisée afin de voir si la classification de Florent Cornuet permettait une meilleure mise en évidence des interactions avec la santé.

### 3.6 Analyse des interactions avec la santé

La première analyse réalisée afin d'étudier les interactions entre la couverture du sol et l'indice de masse corporelle a été de calculer les corrélations linéaires entre tous les paramètres. Les tables contenant toutes les données ont donc été importées dans le logiciel R afin de calculer des matrices de corrélations linéaires et de représenter les résultats sous forme d'une matrice colorée, avec la fonction *levelplot*.

Des régressions linéaires univariées ont ensuite été calculées en utilisant l'indice de masse corporelle comme variable dépendante et une donnée de couverture du sol comme variable explicative. Ces régressions ont été calculées dans Matlab avec la fonction *polyfit* et des graphiques (*scatterplots*) ont été créés pour représenter les résultats. Un test a également été réalisé afin de calculer des régressions linéaires géographiquement pondérées avec le logiciel GWR4 développé par le centre de géoinformatique de l'université de St Andrews <http://www.st-andrews.ac.uk/geoinformatics/gwr/gwr-software/>.

Enfin, une analyse a été réalisée avec le logiciel OpenGeoDa, développé par le GeoDa Center. Ce logiciel a été utilisé afin de créer des graphiques bivariés d'autocorrélation spatiale locale (*LISA bivariés - Local Indicator of Spatial Association*). L'autocorrélation spatiale permet de mettre en évidence la similarité entre la valeur d'un point et celle des points voisins. Dans le cas d'une analyse bivariée, le but est de mettre en évidence des corrélations entre la valeur d'une variable (par exemple l'IMC) pour un point donné et la valeur d'une autre variable (par exemple le pourcentage de végétation) pour les points voisins. On classe ainsi tous les points dans cinq catégories :

- **Low-Low** : Points possédant une valeur faible pour la variable 1 et entourés de points possédant une valeur faible pour la variable 2.
- **Low-High** : Points possédant une valeur faible pour la variable 1 et entourés de points possédant une valeur élevée pour la variable 2.
- **High-Low** : Points possédant une valeur élevée pour la variable 1 et entourés de points possédant une valeur faible pour la variable 2.
- **High-High** : Points possédant une valeur élevée pour la variable 1 et entourés de points possédant une valeur élevée pour la variable 2.
- **Non significatif** : Points pour lesquels aucune relation significative n'est observée entre la valeur de la variable 1 pour ce point et la valeur de la variable 2 pour les points voisins.

Afin de calculer ces classes, la distance pour laquelle le voisinage est considéré doit être définie. Un fichier de poids est ensuite créé pour ce voisinage, permettant de donner un poids plus faible aux points plus éloignés. Les points situés à une distance supérieure à la distance définie pour le voisinage obtiennent un poids nul et ne sont donc pas considérés pour le calcul de la corrélation. Pour ce projet, une distance de 1800 m a été choisie car une étude précédente avait montré que cette distance était appropriée.

Les résultats sont représentés sur une carte en assignant une couleur à chaque point en fonction de la catégorie à laquelle le point appartient. Une carte de significativité est également créée, permettant de voir si les corrélations obtenues sont significatives et peuvent être considérées comme représentatives de la situation. Cette significativité est calculée en réalisant des permutations aléatoires des variables des différents points, puis en calculant à nouveau les résultats des corrélations. Si les corrélations obtenues avec les données réelles sont proches de celles obtenues avec des permutations aléatoires, le résultat n'est pas significatif.



## 4 Résultats des classifications et discussion

### 4.1 Classification avec Erdas Imagine

La figure 4.1 présente la réflectance moyenne des différentes classes pour les quatre bandes de l'image multispectrale, selon les pixels donnés comme référence pour la classification. On constate donc que sur les bandes 1 (G) et 2 (R), les réflectances des différentes classes sont proches, particulièrement pour le sol et les zones urbaines et pour les forêts et la végétation. En revanche, la bande 3 (NIR) semble bien adaptée pour discriminer les classes. Celles-ci sont également distinctes sur la bande 4 (MIR), sauf pour la végétation et les forêts. La combinaison des quatre bandes semble donc permettre de différencier les cinq classes.

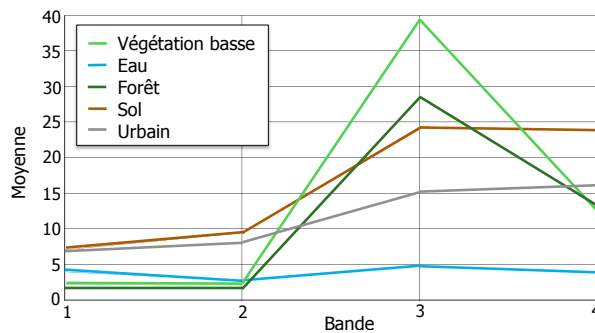


FIGURE 4.1 – Réflectance moyenne des différentes classes sur les quatre bandes pour les échantillons donnés comme référence (réflectances données par Erdas).

La figure 4.2 présente le résultat final de la classification (meilleur résultat, obtenu avec la méthode *maximum likelihood*). Une image en plus grande taille est disponible dans l'annexe D.1.

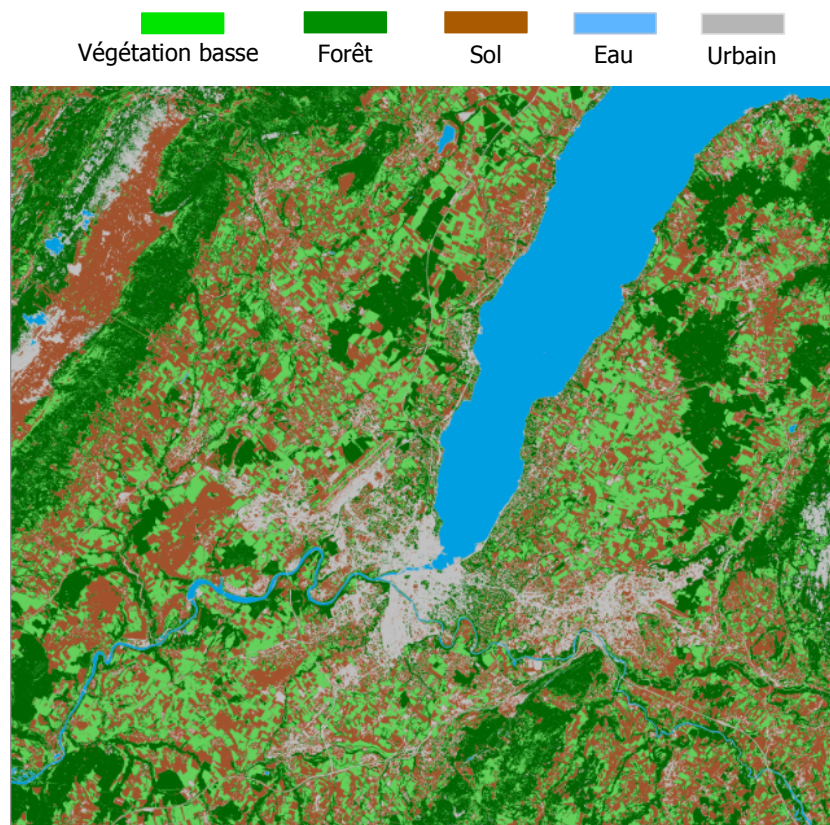


FIGURE 4.2 – Résultat de la classification réalisée avec Erdas Imagine (résolution : 10m).



De manière générale, dans les zones rurales, le sol, la végétation et la forêt sont bien classifiés. Les zones urbaines denses sont également bien classées et les principaux espaces verts à l'intérieur de ces zones sont généralement bien identifiés. On remarque par exemple sur la figure 4.3 que dans le centre urbain dense, la majorité des espaces verts ont été correctement classés. Sur la figure 4.4 on remarque également que le passage des zones urbaines denses à des zones contenant plus de végétation est bien visible.



FIGURE 4.3 – Petits espaces verts correctement classifiés à l'intérieur du centre urbain dense.



FIGURE 4.4 – Passage visible d'une zone urbaine dense à une zone plus verte.

Différents problèmes peuvent toutefois être observés dans la classification. Tout d'abord, comme on peut le constater sur la figure 4.5, les bordures des cours d'eau sont généralement classées dans la catégorie "urbain". De plus, une rivière a été entièrement classifiée dans la classe urbaine, comme on peut le voir sur la figure 4.6. On constate sur l'image infrarouge que pour cette rivière la réflectance était proche de celle des zones urbaines et très éloignée de celle de l'eau, ce qui explique l'erreur de classification.



FIGURE 4.5 – Bordure des cours d'eau classée comme "urbain".

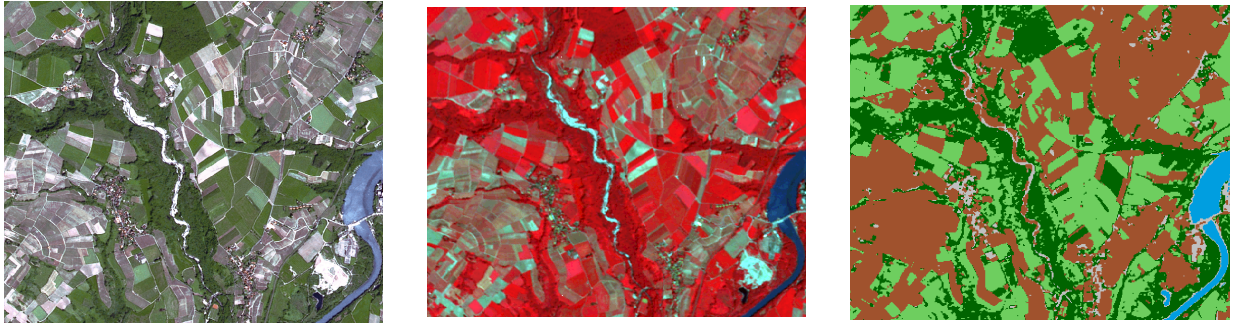


FIGURE 4.6 – Rivière classée comme "urbain".

Un problème plus généralisé peut être observé dans les zones urbaines de faible densité, où de nombreuses surfaces sont classées comme "sol" alors qu'elles devraient principalement être constituées de pixels "urbain" et de pixels "végétation" ou "forêt". La figure 4.7 met en évidence deux zones présentant ce problème. On constate sur l'image infrarouge que le mélange des bâtiment et de la végétation crée une surface avec une réflectance éloignée de celle des deux classes, ce qui est certainement dû à un mélange à l'intérieur même des pixels. Comme la réflectance du sol sur la bande 3 (la bande la plus discriminative) est située entre celle de la végétation et celle des zones urbaines, la réflectance du mélange se retrouve certainement plus proche de la réflectance du sol, ce qui peut expliquer la présence de cette classe dans ces régions.

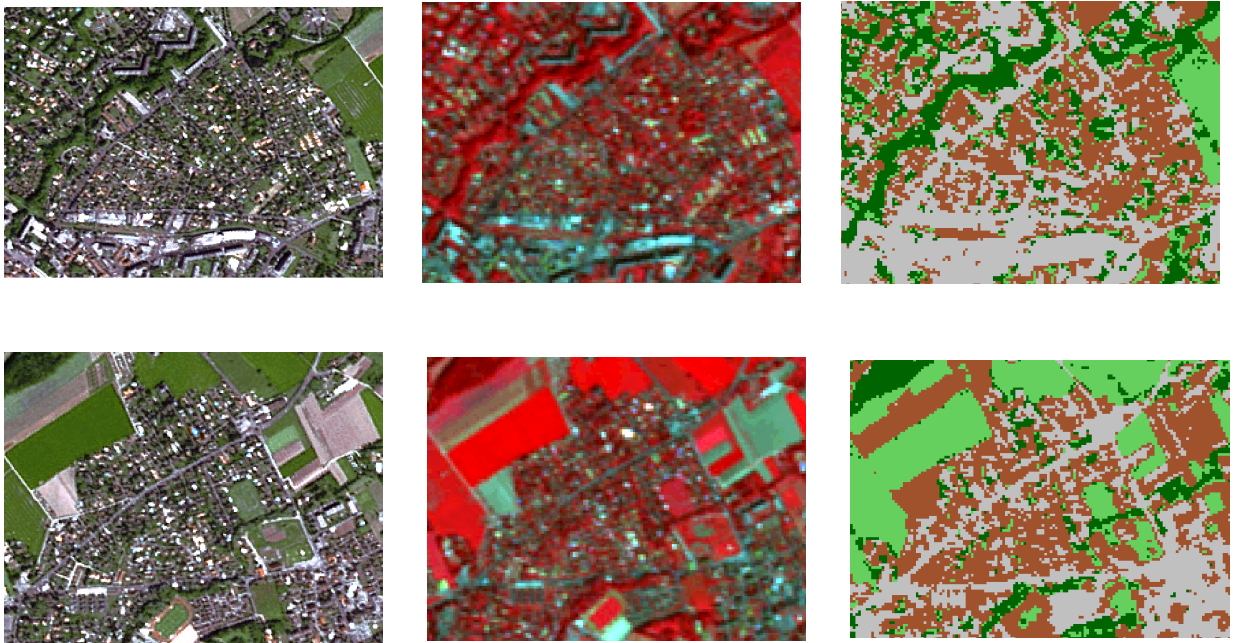


FIGURE 4.7 – Classe "sol" apparaissant abondamment dans les zones urbaines de faible densité

A l'inverse certaines parties de champ très lumineuses se retrouvent dans la classe "urbain", comme on peut le constater sur la figure 4.8.



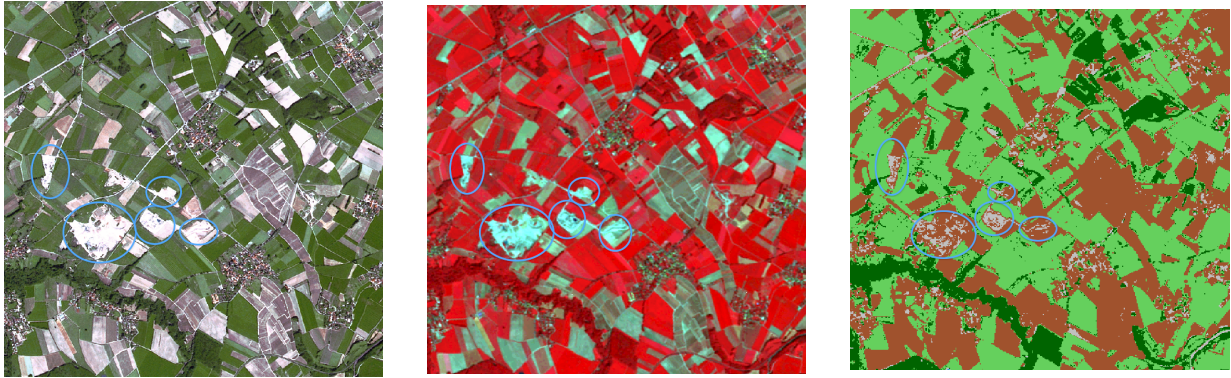


FIGURE 4.8 – Champs très lumineux classés comme zones urbaines.

On peut également observer d'autres problèmes, notamment dans la partie Nord-Ouest de la zone d'étude où des forêts et des nuages ont été classés en zones urbaines et les ombres des nuages ont été classés comme eau. Toutefois, ces problèmes n'influencent pas la zone d'intérêt car cette région est située en dehors du canton de Genève.

## 4.2 Classification avec Matlab

La figure 4.9 présente le meilleur résultat obtenu pour les classifications avec Matlab (méthode *quadratic*). Une image en plus grande taille est disponible dans l'Annexe D.2.

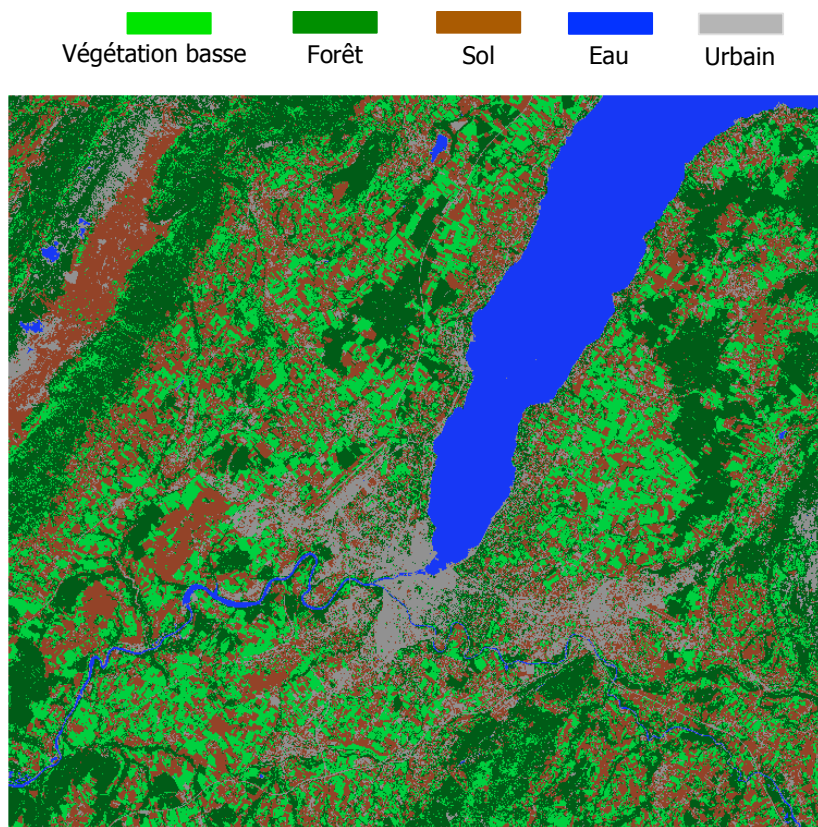


FIGURE 4.9 – Résultat de la classification réalisée avec Matlab.

Ce résultat est très proche de celui obtenu avec Erdas Imagine et il est difficile de mettre en évidence les différences en analysant visuellement les deux images. Ceci s'explique car les méthodes retenues pour les deux classifications sont similaires. Les problèmes identifiés dans la partie précédente s'appliquent donc également ici.

### 4.3 Précisions obtenues

Les polygones manuellement digitalisés sur l'image panchromatique à 5m de résolution et utilisés pour estimer la précision sont présentés sur la figure 4.10. Ces polygones représentent 283'952 pixels répartis entre les différentes classes selon le tableau présentés sur la figure 4.10.

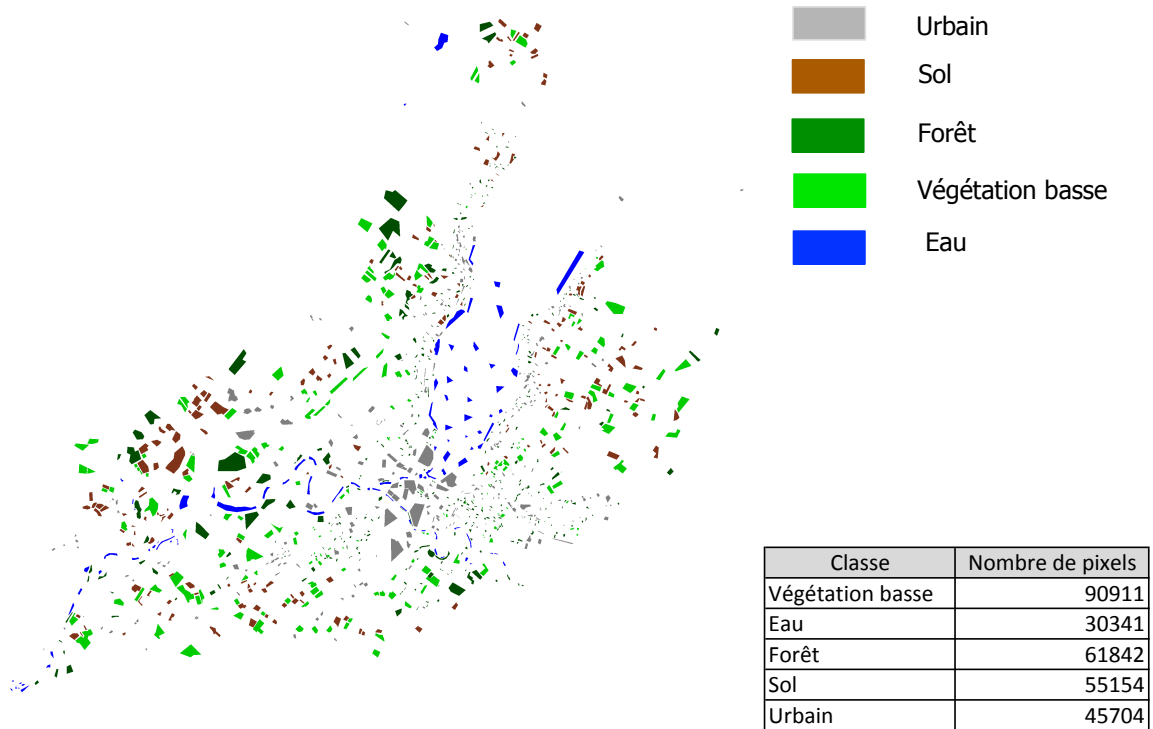


FIGURE 4.10 – Polygones manuellement classés utilisés pour estimer la précision.

La matrice de confusion et les indicateurs de précisions obtenus pour les deux classifications (avec Erdas et avec Matlab) sont présentés dans le tableau ci-dessous.

Matrice de confusion	Classification Erdas					Classification Matlab						
		Végétation	Eau	Forêt	Sol	Urbain		Végétation	Eau	Forêt	Sol	Urbain
	Végétation	87076	1	2232	291	63	Végétation	86798	1	2075	295	73
Eau	0	29487	0	0	10	Eau	0	29470	0	0	10	
Forêt	2660	5	59025	4	234	Forêt	2953	5	59278	4	260	
Sol	1171	16	265	52425	5781	Sol	1157	15	248	52539	5713	
Urbain	4	832	320	2434	39616	Urbain	3	850	241	2316	39648	
<b>OA</b>	0.9495					0.9429						
<b>UA</b>	Végétation	Eau	Forêt	Sol	Urbain	Végétation	Eau	Forêt	Sol	Urbain		
<b>PA</b>	0.9711	0.9997	0.9531	0.8788	0.9169	0.9726	0.9997	0.9484	0.8805	0.9208		
<b>Kappa</b>	0.9578	0.9719	0.9544	0.9505	0.8668	0.9548	0.9713	0.9585	0.9526	0.8675		
<b>Kappa</b>	0.9258					0.9263						

FIGURE 4.11 – Précisions des résultats obtenus avec Erdas et Matlab

On constate à nouveau que les résultats des deux classifications sont très proches. La matrice de confusion permet de mettre en évidence les principaux problèmes de la classification. Pour la classe de végétation, les principaux problèmes sont dus à des pixels de végétation classés en forêt, ce qui s'explique par la réflectance proche des deux milieux, et des pixels classés en sol, qui

correspondent certainement aux champs pour lesquels la couverture végétale est peu épaisse. On constate que ces problèmes sont réciproques puisque des pixels de forêt et de sol sont également classés en végétation. Pour la classe d'eau, on remarque que certains pixels ont été classés dans la classe urbaine, ce qui correspond aux problèmes déjà mis en évidence sur les figures 4.5 et 4.6. Mises à part ces situations, l'eau est en général bien identifiée. Enfin, la matrice de confusion permet également de mettre en évidence le problème de classification entre le sol et la classe urbaine, déjà présenté sur la figure 4.7 et des pixels mal classés entre les forêts et les zones urbaines. Ce dernier problème peut faire référence à des zones de forêt au centre des milieux urbains, notamment les limites de ces zones qui ne sont pas toujours très nettes en raison des ombres causées par les bâtiments et les arbres.

Les indicateurs UA et PA résument ces informations. La précision UA est ainsi très bonne pour l'eau, ce qui indique une forte probabilité que ce qui est classé en eau soit vraiment de l'eau. La précision UA la moins bonne est obtenue pour la classe de sol, ce qui s'explique par la présence des zones de sol dans les milieux urbains de faible densité, qui diminue la probabilité que ce qui est classé en sol soit réellement du sol. Ce problème explique également la précision PA plus faible obtenue pour la classe urbaine, puisque la probabilité que les éléments appartenant à la classe urbaine soient identifiés comme tel est également diminuée.

Finalement, on constate que les précisions globales indiquées par OA et  $\kappa$  sont élevées. Il faut toutefois noter que ces précisions sont certainement surestimées en raison de la comparaison avec des polygones manuellement classés. En effet, lors de la digitalisation des polygones, seules des zones où l'attribution à une classe était relativement sûre ont été choisies. Par conséquent, peu de polygones ont été dessinés par exemple à l'intérieur des zones urbaines peu denses où un mélange important entre les pixels de végétation et de bâtiments rendait la digitalisation d'un polygone difficile. Des erreurs plus importantes sont donc certainement présentes dans ces zones, mais il est difficile de les quantifier. De plus, pour la précision de la classification d'Erdas, il n'est pas certain que les polygones ayant servis pour la classification ne se retrouvent pas en partie parmi ceux utilisés pour la comparaison. Deux jeux de polygones différents ont été digitalisés mais il est possible que certains polygones aient été choisis au même endroit. La précision de la classification de Matlab ne présente pas cet inconvénient, puisque les polygones utilisés pour la comparaison sont ceux laissés de côté lors de la classification.

#### 4.4 Comparaison avec la statistique de la superficie

Les résultats obtenus en comparant les classifications avec les données de la statistique Suisse de la superficie sont présentés dans le tableau ci-dessous.

Matrice de confusion	Classification Erdas					Classification Matlab				
	Végétation + Sol	Eau	Forêt	Urbain		Végétation + Sol	Eau	Forêt	Urbain	
Végétation + Sol	11719	24	743	717		Végétation + Sol	11708	29	694	670
Eau	0	5881	1	0		Eau	0	5839	0	1
Forêt	822	58	3757	84		Forêt	844	52	3816	81
Urbain	478	151	56	1714		Urbain	429	160	41	1761
<b>OA</b>	0.8804					0.8851				
<b>UA</b>	Végétation + Sol	Eau	Forêt	Urbain		Végétation + Sol	Eau	Forêt	Urbain	
<b>PA</b>	0.8876	0.9998	0.7958	0.7145		0.8937	0.9998	0.7962	0.7365	
<b>Kappa</b>	0.9001	0.9619	0.8244	0.6815		0.9019	0.9604	0.8385	0.7008	
	0.818					0.8254				

FIGURE 4.12 – Précisions obtenues en comparant les résultats des classifications avec la statistique Suisse de la superficie

Pour rappel, pour cette comparaison, les classes de végétation et de sol ont été agrégées.

On constate que la précision globale (OA et  $\kappa$ ) est moins bonne que celle obtenue en comparant les résultats avec les polygones manuellement digitalisés. A nouveau, la matrice de confusion permet de mettre en évidence les principaux problèmes, qui sont similaires à ceux déjà observés avec les résultats précédents. On remarque en effet à nouveau les erreurs de classification entre les zones de végétation et de forêt, entre la végétation (le sol) et les zones urbaines et entre l'eau et les zones urbaines. En ce qui concerne les forêts et les zones de végétation, il faut toutefois noter que les attributions des catégories de la statistique de la superficie à l'une de ces classes n'étaient pas toujours claires. Par exemple la catégorie "parcs publics" qui a été assimilée à la végétation, pourrait aussi bien comprendre des arbres et donc correspondre à la classe forêt. De même pour les catégories "jardins familiaux" ou "horticulture". Ces incertitudes peuvent donc être source d'erreurs de classification supplémentaires.

De plus, la statistique de la superficie donne la classe pour un point précis et l'on considère ici que la classe de ce point correspond à la classe majoritaire dans le pixel de 10x10m classifié, ce qui n'est pas forcément correct. Il est donc difficile de savoir si cette précision est plus objective que la précédente, mais cette comparaison a l'avantage d'éviter les problèmes de surestimation de la précision dus à la digitalisation manuelle des polygones.

#### 4.5 Comparaison avec la classification de Florent Cornuet

Les classifications ont également été comparées avec la classification réalisée par Florent Cornuet à partir des données LiDAR et les résultats sont présentés dans le tableau ci-dessous.

Matrice de confusion	Classification Erdas					Classification Matlab					
		Végétation + Sol	Eau	Forêt	Urbain		Végétation + Sol	Eau	Forêt	Urbain	
Végétation + Sol		1.43E+07	1.51E+03	3.49E+06	3.39E+06	Végétation + Sol		1.42E+07	1.52E+03	3.39E+06	3.34E+06
Eau		5.40E+03	2.51E+06	7.26E+04	1.04E+05	Eau		5.23E+03	2.51E+06	6.99E+04	1.02E+05
Forêt		1.31E+06	2.67E+02	7.88E+06	5.74E+05	Forêt		1.37E+06	2.64E+02	8.03E+06	6.06E+05
Urbain		2.61E+06	8.51E+04	3.66E+06	1.40E+07	Urbain		2.59E+06	9.15E+04	3.62E+06	1.40E+07
<b>OA</b>	0.7166					0.7187					
<b>UA</b>	Végétation + Sol	Eau	Forêt	Urbain		Végétation + Sol	Eau	Forêt	Urbain		
<b>PA</b>	0.6748	0.9323	0.8071	0.6881		0.679	0.934	0.8024	0.6901		
	0.7846	0.9666	0.5218	0.7751		0.7822	0.9641	0.5314	0.7761		
<b>Kappa</b>	0.5884					0.5916					

FIGURE 4.13 – Précisions obtenues en comparant les résultats des classifications avec la classification réalisée par Florent Cornuet

Il faut noter ici que par rapport aux comparaisons précédentes, le nombre de pixels comparés est bien supérieur puisque tous les pixels de la zone commune aux deux classifications (avec une résolution de 1m) ont pu être comparés.

La précision globale obtenue avec cette comparaison est très faible ( $\kappa < 0.6$ ), ce qui s'explique en grande partie par les différences de résolution. En effet, les classifications réalisées à partir des images SPOT ont une résolution de 10m, alors que la classification de Florent Cornuet a une résolution de 1m. Une grande partie des détails mis en évidence avec la seconde classification ne sont donc pas visibles sur la première, ce qui cause des erreurs de classification. La figure 4.14 montre un exemple de la différence du niveau de détail entre les deux classifications.



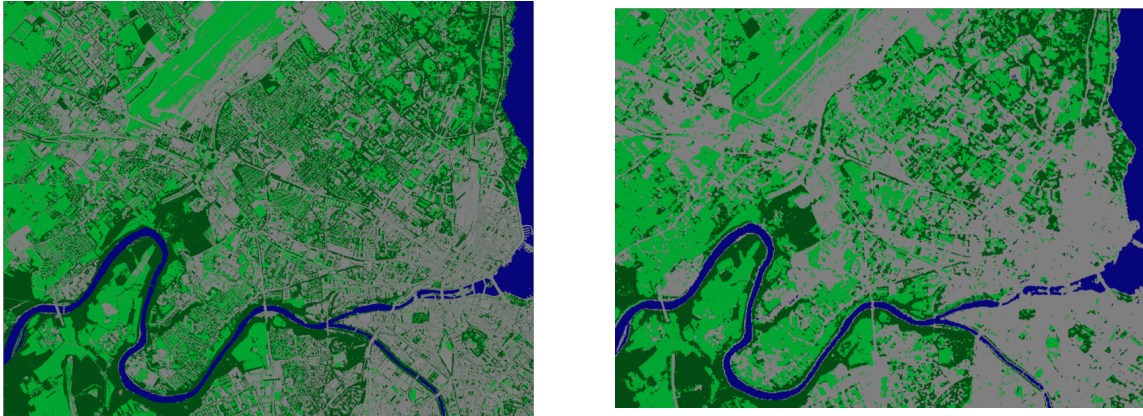


FIGURE 4.14 – Différence du niveau de détail entre les deux classifications

De façon générale, on constate que les principaux problèmes de la classification sont similaires aux problèmes déjà mis en évidence. On remarque toutefois que de nombreux pixels d'eau sont également mal classés alors que cette classe était normalement bien identifiée. Ceci peut s'expliquer notamment par des pixels d'eau isolés dans des zones d'ombre, tel que montré sur la figure 4.15. En comparant tous les pixels, ces fautes peuvent également être mises en évidence.



FIGURE 4.15 – Pixels classés comme eau dans les zones d'ombre.

La classification de Florent Cornuet a également été comparée avec les polygones manuellement digitalisés et avec la statistique Suisse de la superficie. Les résultats sont présentés dans le tableau ci-dessous.

Matrice de confusion	Avec les polygones digitalisés dans Erdas				Avec la Statistique Suisse de la Superficie				
	Végétation + Sol	Eau	Forêt	Urbain	Eau	Perméable	Végétation	Bâtiment	Imperméable
Végétation + Sol	1.43E+06	2.71E+03	2.97E+04	1.75E+05	0	0	0	0	0
Eau	1.50E+02	7.39E+05	2.00E+00	9.00E+01	237	947	21	95	133
Forêt	2.17E+04	6.18E+03	1.30E+06	3.20E+05	9	6	512	5	41
Urbain	1.48E+05	1.62E+04	1.32E+04	1.90E+06	0	71	51	154	164
Imperméable					28	88	27	30	579
OA	0.8798				0.7513				
UA	Végétation + Sol	Eau	Forêt	Urbain	Eau	Perméable	Végétation	Bâtiment	Imperméable
PA	0.8732	0.9997	0.7881	0.9149	1.0000	0.7859	0.9078	0.3242	0.7699
	0.8939	0.9672	0.9679	0.7935	0.7670	0.8516	0.8380	0.5423	0.6314
Kappa	0.8334				0.6697				

FIGURE 4.16 – Précisions obtenues en comparant la classification de Florent Cornuet avec les polygones digitalisés dans Erdas et avec la statistique Suisse de la superficie

La précision relativement faible obtenue avec les polygones manuellement digitalisés est à nouveau due aux problèmes de résolutions en raison desquels même en digitalisant manuellement les polygones sur l'image panchromatique à 5m, des erreurs sont introduites.

Concernant la comparaison avec la statistique Suisse de la superficie, on constate étonnamment que la précision calculée est inférieure à celle obtenue avec les classifications précédentes. Ce résultat indique que la comparaison avec la statistique n'est peut-être pas très adaptée pour estimer la précision des résultats, en raison de la difficulté de trouver des correspondances correctes entre les classes.

En conclusion, cette comparaison permet d'obtenir une précision globale, par rapport à une classification très précise, à 1m de résolution. Au vu de la résolution de l'image classifiée (10m) il n'est pas étonnant que cette précision soit très faible. Pour l'étude des interactions avec la santé, les résultats de la classification de Florent Cornuet seront également utilisés afin de voir si la précision supérieure de cette classification permet de mieux mettre en évidence des interactions.

## 4.6 Conclusion sur les classifications

Après les différentes comparaisons réalisées, il est difficile de conclure sur la précision globale de la classification. La comparaison avec les polygones manuellement digitalisés donne une précision qui tient compte de la résolution de l'image et de la précision que l'on pourrait idéalement atteindre. Par contre, elle présente l'inconvénient d'être certainement surestimée en raison de la digitalisation manuelle des polygones, souvent choisis dans les zones les plus faciles à classifier. La comparaison avec la statistique Suisse de la superficie est plus objective mais il est difficile de définir une correspondance correcte entre les classes et elle suppose que la classe d'un pixel de 10x10 m puisse être comparée avec la classe d'un point échantillonné à l'intérieur de ce pixel. Finalement, la comparaison avec la classification de Florent Cornuet donne une précision plus réelle de la classification, mais ne tient pas compte des différences de résolution en raison desquelles il serait très difficile d'obtenir une bonne précision avec cette comparaison. Ces deux classifications sont également basées sur des méthodes différentes : la classification avec les données LiDAR étant particulièrement utile pour distinguer les bâtiments du sol ou les arbres de la végétation basse, mais s'appliquant moins facilement à de grandes surfaces.

Afin d'obtenir une meilleure estimation de la précision, il faudrait pouvoir comparer les résultats avec une donnée de résolution similaire. Par exemple, une classification en 27 classes a déjà été réalisée par [14] pour le canton de Genève à partir des mêmes données SPOT, combinées avec des données vectorielles (classification CCSA04).

De façon générale, les résultats ont montré que les classes principales pouvaient bien être distinguées, et que les espaces verts pouvaient être identifiés au centre des grandes zones urbaines. Pour atteindre une précision plus grande, plusieurs problèmes de classification devraient être résolus, principalement le problème des zones urbaines classifiées comme sol. Ce problème pourrait influencer les analyses des interactions avec la santé car il peut conduire à une surestimation des zones vertes lorsque les catégories "sol" et "végétation" sont agrégées.

Des images de plus grande résolution pourraient également être utilisées. Dans le cadre de ce projet, il aurait peut-être été plus judicieux de choisir des images de plus grande résolution (orthophotos disponibles par exemple) et si nécessaire de dégrader la résolution pour diminuer le poids des fichiers. Une résolution intermédiaire entre celles des données SPOT et des orthophotos aurait ainsi pu être atteinte et ces données auraient présentés l'avantage d'être plus récentes. En effet, les données utilisées pour ce projet datent de 2004 et il est possible que des changements importants aient déjà eu lieu dans certaines régions du canton.

Pour la suite des analyses, la classification réalisée avec Matlab a été utilisée. Les résultats des deux classifications étant très proches, la classification réalisée avec Erdas aurait aussi pu être choisie. La classification réalisée avec Matlab a été préférée car elle avait une précision très légèrement supérieure, et pour la comparaison avec les polygones manuellement digitalisés, cette précision n'était pas influencée par d'éventuels polygones ayant déjà servis pour la classification.



## 5 Résultats de l'étude des interactions avec la santé et discussion

### 5.1 Corrélations linéaires

Les corrélations linéaires ont été calculées entre les données de couverture du sol et les données de santé. Les résultats mettent en évidence les corrélations existant entre les différents paramètres de couverture du sol (corrélations négatives entre les surfaces liées à la végétation et les surfaces urbaines) ainsi que des corrélations relativement importantes entre le revenu médian et la couverture du sol (pour les adultes, corrélations de -0.61 avec le pourcentage de zones urbaines dans la cellule de 200m, et de +0.64 avec le pourcentage cumulé de forêt, sol et végétation basse dans la même cellule). Les matrices de corrélations sont disponibles dans les Annexes E.1 à E.4.

La figure 5.1 présente les corrélations obtenues entre les données IMC (BMI) et les données de couverture du sol, en utilisant la classification réalisée avec Matlab. Les nombres 200, 800 et 3200 indiquent la taille de la grille utilisée pour le calcul du pourcentage occupé par des différentes classes. Les valeurs numériques des corrélations sont disponibles dans les Annexes E.1 et E.3.

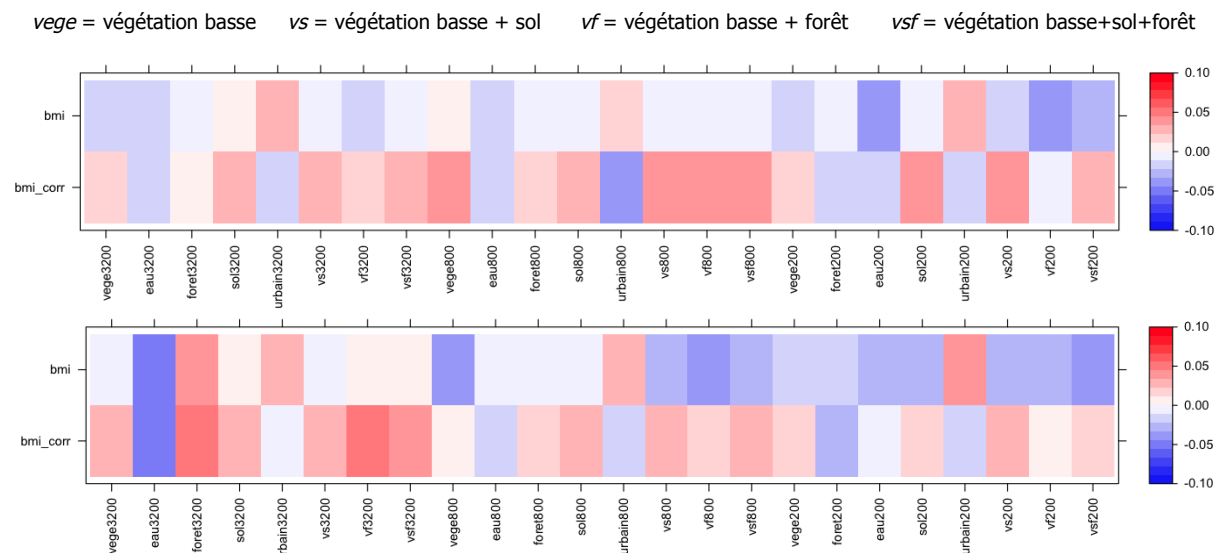


FIGURE 5.1 – Corrélations linéaires entre les données de couverture du sol obtenues avec la classification réalisée avec Matlab, et les données IMC des adultes (haut) et des enfants (bas)

On constate que les corrélations sont toutes très faibles (inférieures à 0.1), ce qui indique qu'il y a très peu de relation linéaire entre les données d'IMC et les données de couverture du sol. De façon générale, on remarque que les corrélations obtenues avec l'IMC corrigé sont plus grandes que celles obtenues avec l'IMC non corrigé, ce qui indique que la valeur corrigée est peut-être plus apte à mettre en évidence les relations entre les données. On remarque également que les corrélations obtenues avec les zones urbaines sont toujours négatives (augmentation de l'IMC correspondant à une diminution du pourcentage de zones urbaines) alors que celles obtenues avec les zones de végétation et de sol (isolément et cumulées) sont toujours positives (augmentation de l'IMC associée à une augmentation des pourcentages de sol et de végétation).

La corrélation positive maximale pour les données concernant les adultes est obtenue entre l'IMC corrigé et le pourcentage cumulé de végétation, sol et forêt dans les cellules de 800m (corrélation de +0.04). La corrélation négative maximale est atteinte entre l'IMC corrigé et le pourcentage de zones urbaines dans les cellules de 800m également (-0.037). Pour les enfants, la corrélation positive maximale est obtenue entre l'IMC corrigé et le pourcentage de forêt dans les cellules de 3200m (+0.055) et la corrélation négative maximale entre l'IMC corrigé et le pour-

centage d'eau dans la cellule de 3200m (-0.051). Il faut noter ici que le pourcentage d'eau est corrélé avec les zones urbaines puisque le centre urbain dense est situé proche du lac.

Ces résultats sembleraient donc indiquer une tendance inverse de celle attendue, soit une augmentation de l'IMC dans les milieux plus naturels. Toutefois, vu les valeurs très faibles des corrélations, ces résultats ne sont certainement pas significatifs et ne permettent pas de conclure sur une relation linéaire entre la couverture du sol et l'indice de masse corporelle.

La figure 5.2 présente les corrélations obtenues en utilisant la classification réalisée par Florent Cornuet. Les valeurs numériques sont disponibles dans les Annexes E.2 et E.4

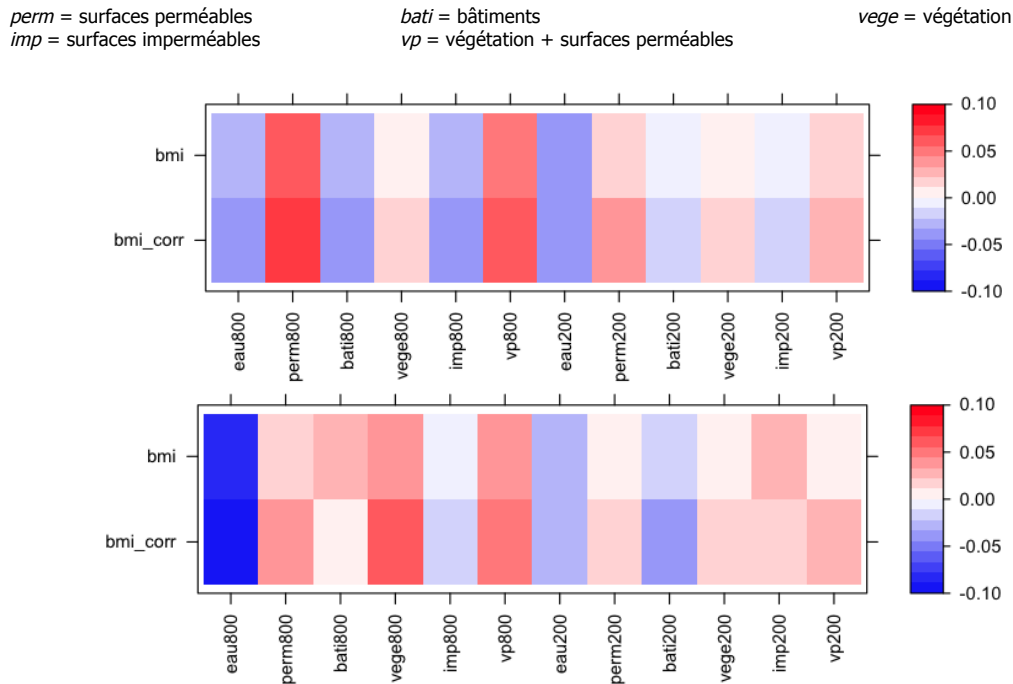


FIGURE 5.2 – Corrélations linéaires entre les données de couverture du sol obtenues avec la classification réalisée par Florent Cornuet, et les données IMC des adultes (haut) et des enfants (bas)

A nouveau on constate qu'aucune corrélation ne dépasse 0.1, mais de manière générale, les corrélations sont plus élevées que celles observées précédemment. Pour les adultes, les corrélations sont toujours du même signe pour l'IMC et l'IMC corrigé. Elles sont négatives pour les classes eau, surfaces imperméables et bâtiments et positives pour les surfaces perméables et la végétation. La tendance est moins claire pour les enfants où les bâtiments et les surfaces imperméables obtiennent parfois des corrélations positives (mais très proches de 0).

La corrélation positive maximale obtenue pour les adultes correspond à l'IMC corrigé et les surfaces perméables dans les cellules de 800m (+0.07) et la corrélation négative maximale est obtenue entre l'IMC corrigé et les surfaces imperméables dans les cellules de 800m également (-0.044). Pour les enfants, la corrélation positive maximale est observée entre l'IMC corrigé et la végétation, dans les cellules de 800m (+0.057) et la corrélation négative maximale entre l'IMC corrigé et le pourcentage d'eau, toujours dans les cellules de 800m (-0.091).

Ces résultats indiquent donc les mêmes tendances que celles observées avec la classification réalisée avec Matlab, bien que les précisions des deux classifications et les zones couvertes soient différentes. De plus, on constate que les corrélations maximales sont toujours observées avec

les cellules de 800m pour la classification réalisée par Florent Cornuet ou 800m et 3200m pour la classification réalisée avec Matlab (pour rappel, les résultats de la classification de Florent Cornuet n'ont pas été utilisés avec les cellules de 3200m). On pourrait donc déduire que les cellules de 200m semblent moins appropriées pour mettre en évidence les corrélations. Toutefois, il faut à nouveau noter que les corrélations observées sont très faibles et ne permettent pas de donner des conclusions significatives.

## 5.2 Régressions linéaires

Des régressions linéaires ont également été calculées, en utilisant l'IMC corrigé comme variable dépendante et une donnée de couverture du sol comme variable explicative. La figure 5.3 présente la répartition de l'IMC des adultes en fonction des différentes variables de couverture du sol, pour les cellules de 800m.

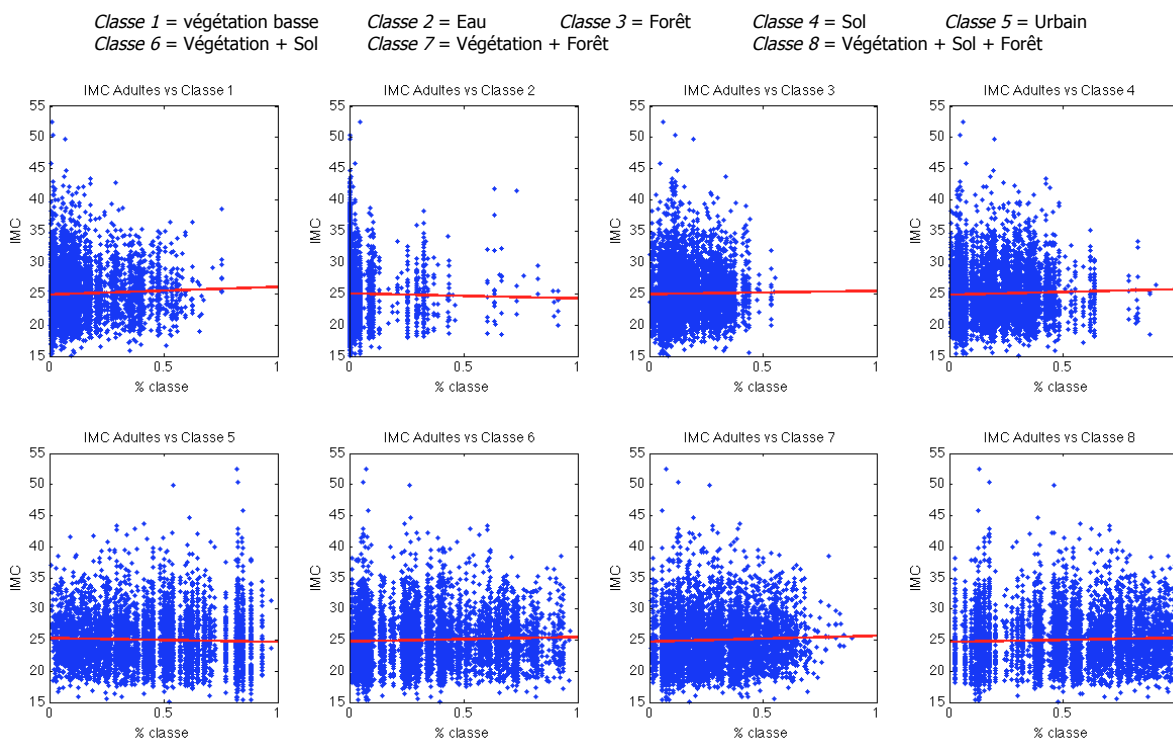


FIGURE 5.3 – Régressions linéaires entre l'IMC corrigé et les pourcentages représentés par les classes de couverture du sol dans les cellules de 800m

On constate donc qu'aucune relation linéaire n'est visible entre l'IMC et les données de couverture du sol. Pour cette raison, les résidus obtenus avec des régressions linéaires sont très grands et les régressions ne sont pas significatives. Des graphiques très similaires sont obtenus avec les autres tailles de cellules, avec les données des enfants et avec la classification de Florent Cornuet. La figure 5.4 présente un exemple obtenu pour les enfants, avec la classification de Florent Cornuet et les cellules de 800m. Le nombre de données inférieur explique l'aspect moins continu du graphique.

Des tests ont également été réalisés pour calculer des régressions linéaires géographiquement pondérées sur les jeux de données, avec le logiciel GWR4. Ces calculs étaient toutefois très lourds et pour les cas traités, aucun résultat significatif n'a pu être obtenu. L'analyse n'a donc pas été poursuivie.

*Classe 1 = Eau*                      *Classe 2 = Surfaces perméables*                      *Classe 3 = Bâtiment*  
*Classe 4 = Végétation*                      *Classe 5 = Surfaces imperméables*                      *Classe 6 = Végétation + Surfaces perméables*

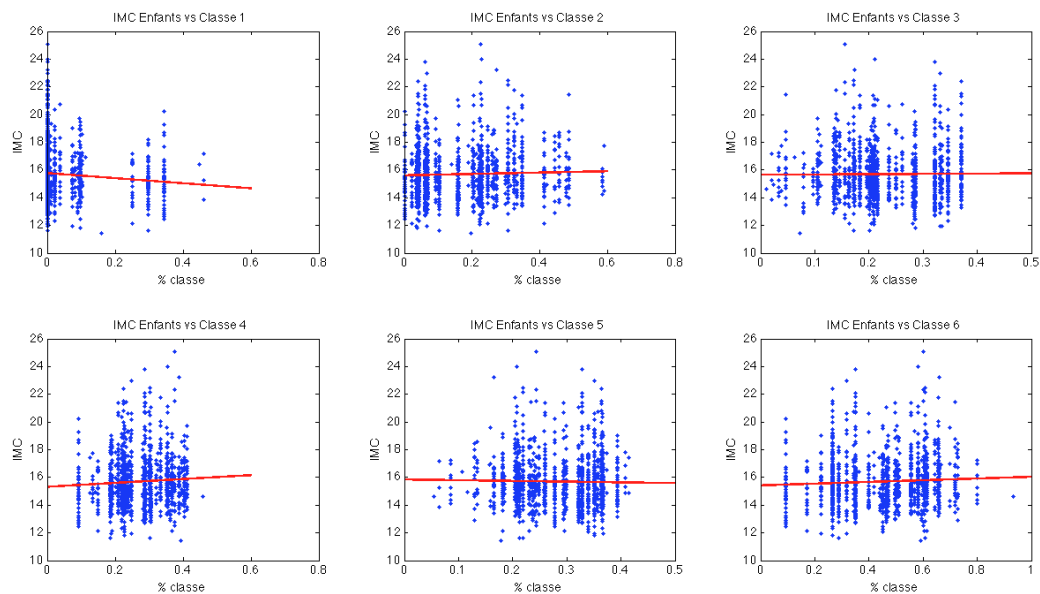


FIGURE 5.4 – Régressions linéaires entre l’IMC corrigé et les pourcentages représentés par les classes de couverture du sol dans les cellules de 800m, avec les données de la classification de Florent Cornuet

### 5.3 LISA bivariés

Finalement, des graphiques LISA bivariés ont été réalisés entre les couples de variables, afin de voir si des variations spatiales et des relations localisées pouvaient être mises en évidence. Les figures 5.5 à 5.10 donnent différents exemples de résultats obtenus pour l’IMC corrigé et les classes de couverture du sol liées à la végétation. Pour ces exemples, 999 permutations ont été utilisées pour tester la significativité et la distance du voisinage est de 1800 m.

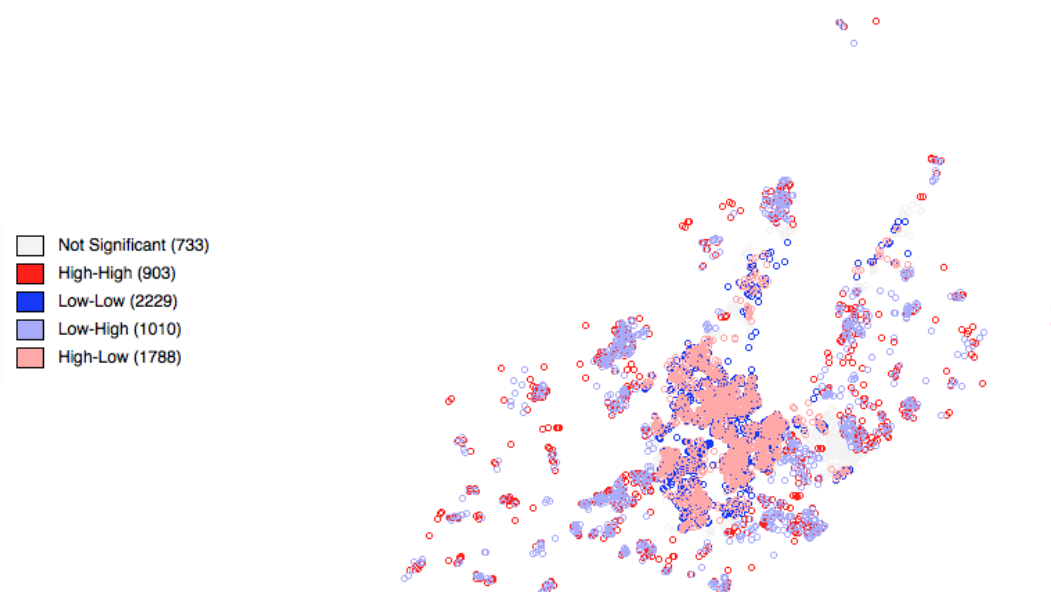


FIGURE 5.5 – LISA bivariés entre le BMI corrigé des adultes et le pourcentage cumulé de végétation et forêt dans les cellules de 3200m, avec la classification réalisée avec Matlab.

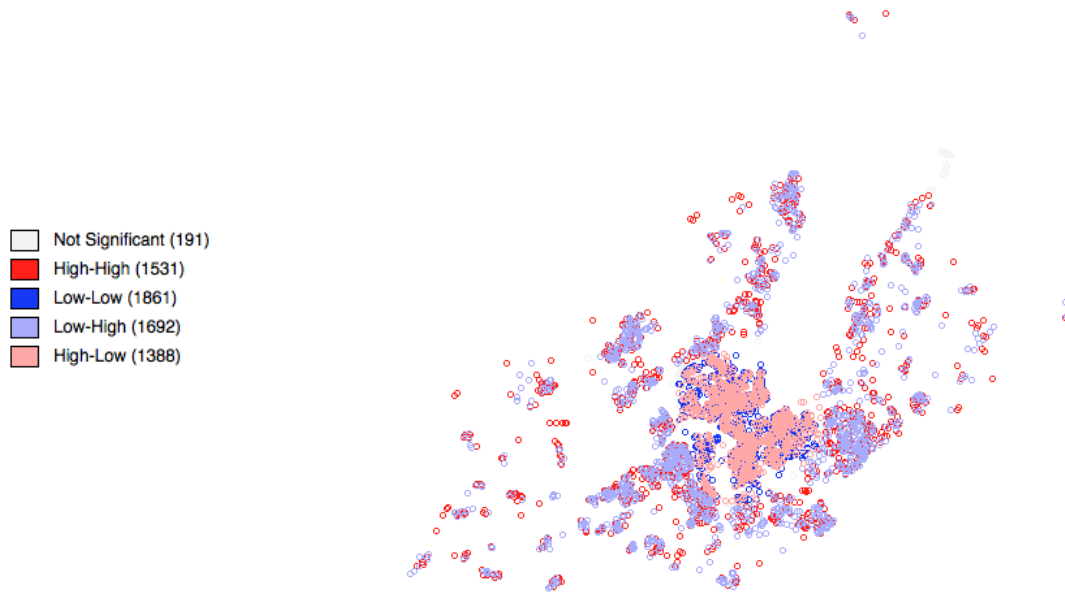


FIGURE 5.6 – LISA bivariés entre le BMI corrigé des adultes et le pourcentage cumulé de végétation, sol et forêt dans les cellules de 800m, avec la classification réalisée avec Matlab.

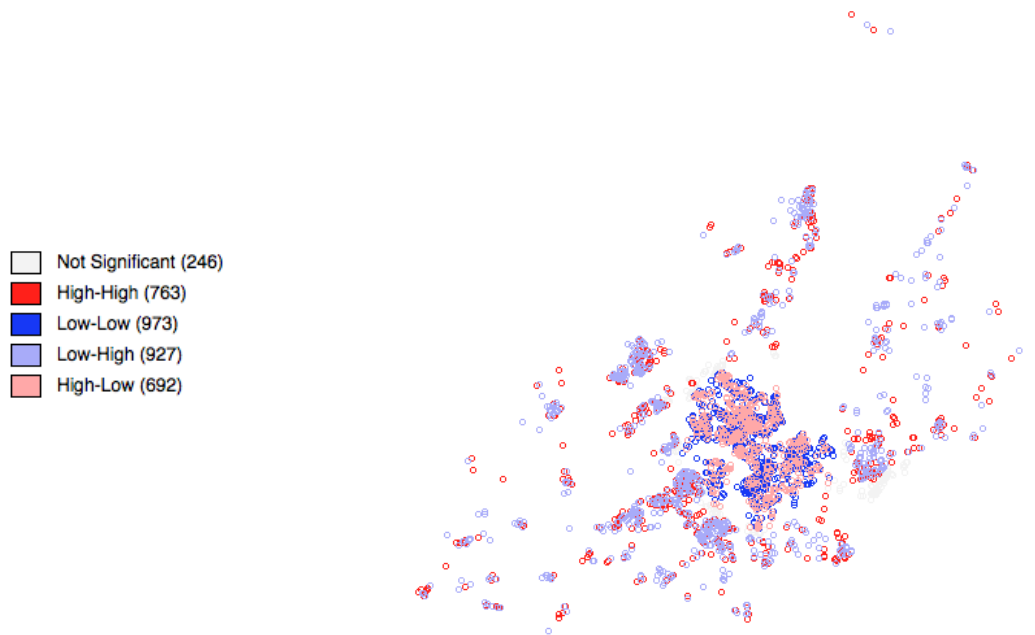


FIGURE 5.7 – LISA bivariés entre le BMI corrigé des enfants et le pourcentage cumulé de surfaces de végétation, de sol et de forêt dans les cellules de 200m, avec la classification réalisée avec Matlab.

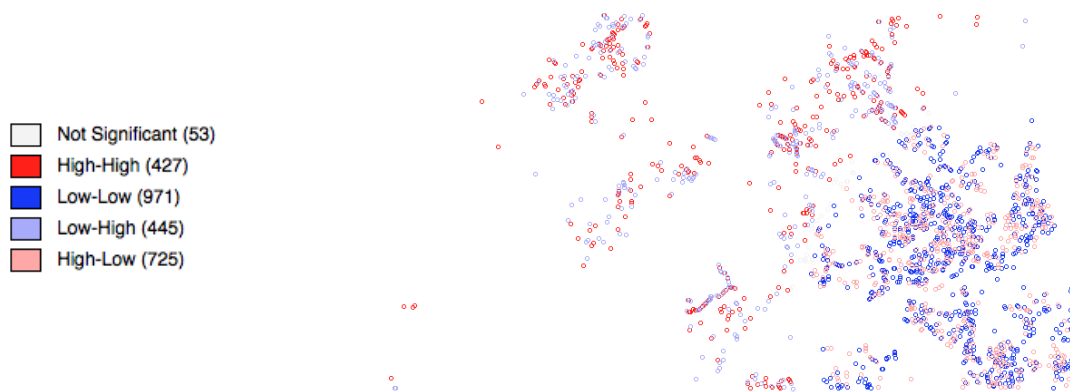


FIGURE 5.8 – LISA bivariés entre le BMI corrigé des adultes et le pourcentage de surfaces perméables dans les cellules de 800m, avec la classification réalisée par Florent Cornuet.

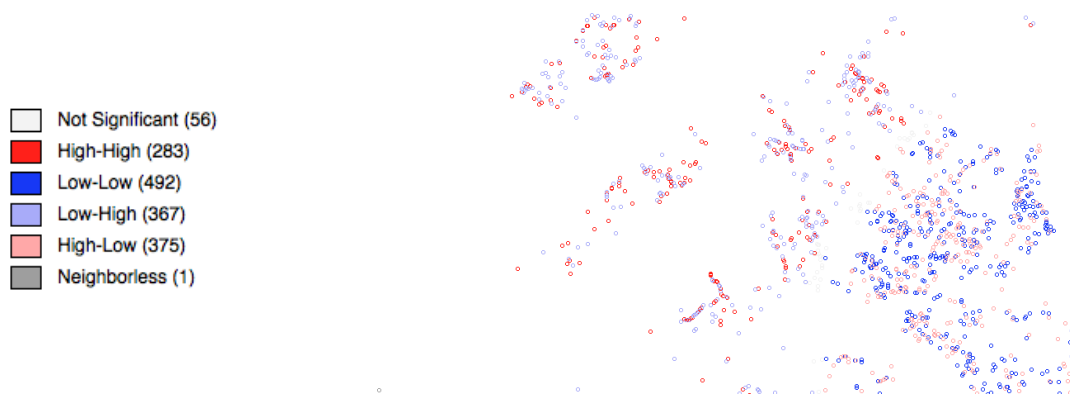


FIGURE 5.9 – LISA bivariés entre le BMI corrigé des enfants et le pourcentage cumulé de surfaces perméables et de végétation dans les cellules de 800m, avec la classification réalisée par Florent Cornuet.

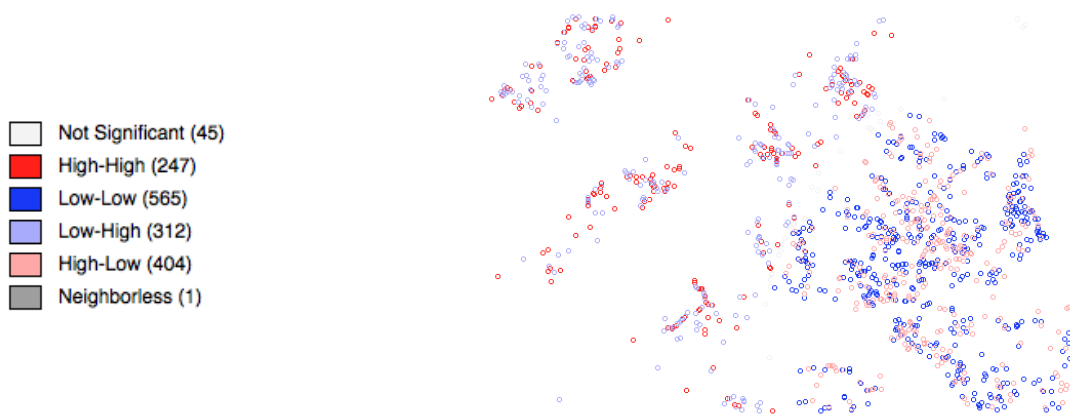


FIGURE 5.10 – LISA bivariés entre le BMI corrigé des enfants et le pourcentage de surfaces perméables dans les cellules de 200m, avec la classification réalisée par Florent Cornuet.

Les résultats obtenus sont similaires entre les différentes classes, avec les différentes classifications et pour les adultes et les enfants. Lorsque les graphiques sont réalisés entre l'IMC et des variables de végétation, deux groupes ressortent toujours : un groupe constitué d'un mélange d'individus Low-Low et Low-High proche du centre urbain et un groupe constitué d'un mélange d'individus High-High et High-Low en périphérie. Dans chaque groupe, le nombre d'individus répartis dans les deux catégories est similaire. Ces deux groupes n'indiquent pas une relation entre l'IMC et la couverture du sol mais mettent en évidence la répartition spatiale des classes de végétation. En effet, proche du centre, les pourcentages de végétation sont faibles, mais on trouve aussi bien des individus avec des IMC faibles qu'avec des IMC élevés, on obtient donc des associations Low-Low et Low-High. En périphérie, les pourcentages de classes liées à la végétation sont plus élevés, mais les IMC sont à nouveau élevés ou bas, et on retrouve donc des associations High-Low et High-High. Lorsque les graphiques sont réalisés entre l'IMC et la classe urbaine, les groupes sont inversés.

Ces observations ne permettent donc pas de conclure sur une relation entre l'IMC et la couverture du sol.

#### 5.4 Conclusion sur les interactions avec la santé

Les résultats n'ont donc pas permis de mettre en évidence une relation entre l'indice de masse corporelle et la couverture du sol. Les corrélations linéaires entre les paramètres permettent de mettre en évidence une tendance inverse à celle attendue, soit une augmentation de l'IMC lorsque le pourcentage de surfaces urbaines diminue. Toutefois, les valeurs des corrélations sont très faibles et donc certainement non-significatives. De plus, les régressions linéaires et les graphiques LISA bivariés n'ont pas permis de montrer l'existence d'une relation entre les paramètres de couverture du sol et l'indice de masse corporelle.

Une relation existe peut-être entre les données, mais la méthode utilisée n'a pas permis de la mettre en évidence. Il est possible que la précision de la classification soit insuffisante ou que le calcul des pourcentages par cellules ne soit pas adapté, puisque cette méthode n'assure pas que la distance considérée soit la même pour tous les individus (si un individu est situé dans le bord Est de la cellule de 3200m, la couverture du sol prise en compte n'est pas répartie uniformément autour de l'individu, mais est située dans les 3200m en direction de l'Ouest).

Il est également possible qu'aucune relation significative entre l'IMC et la couverture du sol ne soit présente pour les données considérées. Toutefois, les résultats obtenus ne permettent pas d'affirmer que c'est le cas et une analyse plus approfondie devrait être réalisée, par exemple en distinguant les espaces fortement urbanisés des espaces plus ruraux, car ces milieux peuvent présenter des relations différentes [8, 4]. D'autres paramètres peuvent également influencer l'analyse et devraient peut-être être considérés, notamment le genre [6, 3] et l'âge des individus [4] (par exemple la distinction entre adultes actifs et personnes âgées [8]) et la facilité d'accès aux milieux naturels. Enfin, il serait peut-être intéressant de réaliser l'analyse uniquement avec les valeurs d'IMC sortant de la marge normale (IMC compris entre 18.5 et 25 [21]).

## 6 Conclusion

Ce projet a permis de réaliser une classification de la couverture du sol dans le canton de Genève à partir d'images satellites multispectrales SPOT de 10 m de résolution datant de 2004. La classification a été réalisée avec Erdas Imagine et avec Matlab et les meilleurs résultats obtenus avec les deux logiciels sont très proches. La précision de la classification a été estimée en utilisant comme référence des polygones manuellement digitalisés sur les images panchromatiques SPOT à 5 m de résolution. La précision obtenue est bonne ( $\kappa=0.92$ ) mais certainement surestimée en raison de la digitalisation manuelle des polygones, souvent réalisée dans les zones plus évidentes à classer. La classification a également été comparée avec la statistique Suisse de la superficie et la précision finale obtenue est de  $\kappa = 0.82$ . Il est toutefois difficile d'estimer la fiabilité de cette comparaison en raison de la difficulté de trouver une correspondance correcte entre les classes. Finalement, la classification a été comparée avec une classification réalisée par Florent Cornuet à partir de données LiDAR à 1m de résolution. La précision obtenue est faible ( $\kappa=0.59$ ) ce qui s'explique en partie par la différence de résolution des deux classifications. Cette comparaison est toutefois plus fiable que les deux précédentes. Il est donc difficile de conclure sur la précision de la classification mais de façon générale, différents problèmes ont pu être mis en évidence et devraient être résolus pour obtenir une plus grande précision. Les principaux problèmes concernent des éléments urbains assimilés à du sol nu dans les zones urbaines de faible densité, des mélanges entre les classes de végétation basse et de forêt et quelques cas de surfaces d'eau et de champs classés en zone urbaine.

Les résultats obtenus ont ensuite été utilisés afin d'analyser l'interaction entre la couverture du sol et l'indice de masse corporelle (IMC) et de voir si une relation pouvait être mise en évidence. Diverses études ont en effet montré que la présence de zones vertes en milieu urbain était bénéfique pour la santé physique ([1] à [5]) et mentale ([1, 3, 6]) des habitants. De plus, certaines études ont montré que les surfaces vertes pouvaient permettre de favoriser l'activité physique et d'avoir ainsi un impact sur le poids des individus ([4, 5]). L'analyse a été basée sur les données d'IMC de 6663 adultes et 3601 enfants domiciliés dans le canton de Genève. Les pourcentages représentés par chaque classe de couverture du sol ont été calculés dans des cellules de 200, 800 et 3200m et les informations ont été mises en relation avec les données d'IMC des individus domiciliés dans ces cellules. Les interactions entre l'IMC et la couverture du sol ont ensuite été testées à l'aide de coefficients de corrélations linéaires, de calcul de régressions linéaires et de graphique LISA bivariés. Ces trois méthodes n'ont pas permis de mettre en évidence des relations significatives entre la couverture du sol et l'indice de masse corporelle. Une relation entre les variables n'est toutefois pas exclue, mais les méthodes devraient être approfondies afin de mettre en évidence une éventuelle interaction entre la couverture du sol et la santé dans la zone d'étude considérée.



## Références

- [1] P.P.Groenewegen P.Spreeuwenberg. S.de Vries, R.A.Verheij. Natural environments - healthy environments? An exploratory analysis of the relationship between greenspace and health. *Environment and Planning A.*, 35 :1717 – 1731, 2003.
- [2] F.C.Curriero W.F.Stewart B.S.Schwartz J.Feng, T.A.Glass. The built environment and obesity : A systematic review of the epidemiologic evidence. *Health and Place*, (16) :175–190, 2010.
- [3] R.Maheswaran A.C.K.Lee. The health benefits of urban green spaces : a review of the evidence. *Journal of Public Health*, 33(2) :212–222, 2010.
- [4] J.Wolch M.Jerrett K.D.Reynolds G.F.Dunton, J.Kaplan. Physical environmental correlates of childhood obesity : a systematic review. *International Association for the Study of Obesity - Obesity Reviews*, (10) :393–402, 2009.
- [5] M.Hillsdon E.Coombes, A.P.Jones. The relationship of physical activity and overweight to objectively measured green space accessibility and use. *Social Science and Medicine*, 70 :816–822, 2010.
- [6] K.N.Irvine R.A.Fuller L.E.Keniger, K.J.Gaston. What are the benefits of Interacting with Nature? *International Journal of Environmental Research and Public Health*, 10 :913 – 935, 2013.
- [7] C.D.Powell D.A.McNeil R.D.Jacobson L.McLaren M.L.Potestio, A.B.Patel. Is there an association between spatial access to parks/green space and childhood overweight/obesity in calgary, canada? *International Journal of Behavioral Nutrition and Physical Activity*, 6(77), 2009.
- [8] P.P.Groenewegen S.de Vries P.Spreeuwenberg J.Maas, R.A :Verheij. Green space, urbanity, and health : how strong is the relation? *Journal of Epidemiology and Community Health*, 60(7) :587–592, 2006.
- [9] F.Popham R.Mitchell. Greenspace, urbanity and health : relationships in england. *Journal of Epidemiology and Community Health*, 61(8) :681–683, 2007.
- [10] Florent Cornuet. Contribution du modèle lidar pour la caractérisation de surfaces urbaines et le paramétrage de modèles de simulation des flux atmosphériques en pollution de l'air. *Projet de master LASIG -EPFL*, 2011.
- [11] Wikipedia. SPOT (satellite). consulté le 23.11.13. URL : [http://fr.wikipedia.org/wiki/SPOT\\_\(satellite\)](http://fr.wikipedia.org/wiki/SPOT_(satellite)).
- [12] ASTRIUM and EADS Company. Images satellites SPOT. consulté le 23.11.13. URL : <http://www.astrium-geo.com/fr/233-images-satellites-spot>.
- [13] NPOC Swisstopo. Spot5 Radcor Mosaic of Switzerland. 2009.
- [14] J-M.Jaquet S.Schwarzer, K.Allenbach. Cartographie semi-automatisée et multi-sources de la couverture du sol - 1. Périmètre de l'agglomération franco-valdo genevoise (CCSA04 moyenne résolution). *Archives des Sciences*, (62) :107–124, 2009.
- [15] H.Fournier T.Produit. La statistique suisse de superficie - Spatial Analysis. *Master I - SSIE EPFL*, 2007.
- [16] S.Beer-Borst. L'épidémiologie nutritionnelle en Suisse : Un regard sur Genève vaut la peine. 2007. URL : [http://www.nutrinet.ch/fileadmin/documents/nutrinet\\_tagungen/herbsttagung\\_2007/referat\\_beer.pdf](http://www.nutrinet.ch/fileadmin/documents/nutrinet_tagungen/herbsttagung_2007/referat_beer.pdf).

- [17] Office Fédéral de la Santé Publique OFSP. Description des projets de monitoring alimentation et activité physique - bus santé. consulté le 12.11.13. URL : [http://www.bag.admin.ch/themen/ernaehrung\\_bewegung/05190/05294/12331/index.html?lang=fr](http://www.bag.admin.ch/themen/ernaehrung_bewegung/05190/05294/12331/index.html?lang=fr).
- [18] Intergraph. Erdas Imagine - Complete Remote Sensing and Image Analysis. consulté le 04.12.13. URL : <http://geospatial.intergraph.com/products/ERDASIMAGINE/ERDASIMAGINE/Details.aspx>.
- [19] ERDAS-Intergraph. ERDAS Field Guide - Fourth Edition, Revised and Expanded. pages 250–253, 1997. URL : <http://www.ltrs.uri.edu/teaching/FieldGuide.pdf>.
- [20] MathWorks Documentation Center. Classify. consulté le 26.12.2013. URL : <http://www.mathworks.ch/help/stats/classify.html>.
- [21] Wikipedia. Indice de masse corporelle. consulté le 06.01.14. URL : [http://fr.wikipedia.org/wiki/Indice\\_de\\_masse\\_corporelle](http://fr.wikipedia.org/wiki/Indice_de_masse_corporelle).

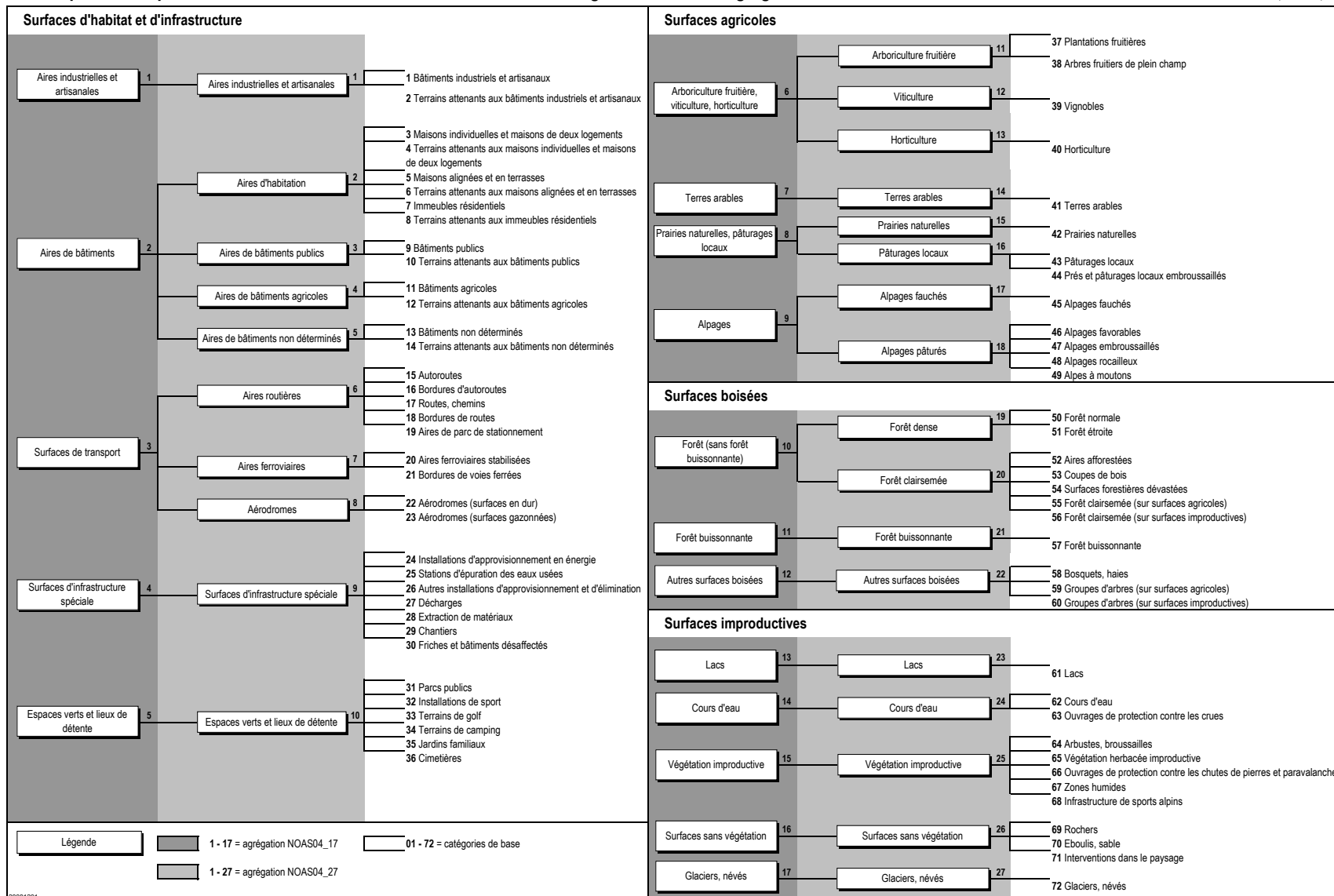


# A Annexe : Catégories de la statistique Suisse de la superficie

## Statistique de la superficie Suisse — Nomenclature standard NOAS04: Catégories de base et agrégations

OFS, Statistique de la superficie

VIII



## B Annexe : Codes Matlab

### B.1 Code Matlabs utilisé pour la classification

```
%Classification Matlab
% =====
% Création de la carte de référence avec les polygones créés dans Erdas
agri_true=imread('agri_true.tif');
eau_true=imread('eau_true.tif');
foret_true=imread('foret_true.tif');
urbain_true=imread('urbain_true.tif');
sol_true=imread('sol_true.tif');

% classes: 0 = rien, 1 = végétation, 2 = eau, 3 = forets, 4 = sol, 5= urbain
agri_true(agri_true>0)=1;
eau_true(eau_true>0)=2;
foret_true(foret_true>0)=3;
sol_true(sol_true>0)=4;
urbain_true(urbain_true>0)=5;
GT=(agri_true)+(eau_true)+(foret_true)+(sol_true)+(urbain_true);

%=====
% Classification
Imm = imread('geneva_ir_cut.tif');

Sz=size(Imm);
Imm_r = reshape(double(Imm), Sz(1)*Sz(2),Sz(3));
GT_r = reshape(GT, Sz(1)*Sz(2),1);

GT_ID = find(GT_r>0);
trainID = randsample(GT_ID, ceil(length(GT_ID)/5));
testID = setdiff(GT_ID, trainID);

[Pred, Err, Posterior] =
classify(Imm_r,Imm_r(trainID,:),GT_r(trainID:),'quadratic');
% Err = apparent error rate: percentage of observations in the TRAINING that
are misclassified
Map_pred=reshape(Pred,Sz(1),Sz(2));

% =====
% Création du fichier .csv à importer dans Quantum GIS pour créer l'image .tif
class=Map_pred;
res = 10; % résolution spatiale de l'image classifiée en [m]
% coordonnées de la zone d'étude dans le système CH1903+
xmin = 484200+res/2;
xmax = 516200-res/2;
ymin = 107700+res/2;
ymax = 136500-res/2;

x=zeros(size(class,1)*size(class,2),1);
y=zeros(size(class,1)*size(class,2),1);
class=zeros(size(class,1)*size(class,2),1);
for j=1:size(class,2)
    for i=1:size(class,1)
        x((j-1)*size(class,1)+i)=xmin+(j-1)*res;
        y((j-1)*size(class,1)+i)=ymax-(i-1)*res;
        class1((j-1)*size(class,1)+i)=class(i,j);
    end
end
matlab = [x y class];
dlmwrite('class_matlab_table.csv',matlab,'precision','%0f')
```



## B.2 Code Matlab utilisé pour les calculs des précisions

```
n=5; % nombre de classes

% Importation et affichage de la classification d'Erdas
class=imread('class5_maxlike.tif');

% Importation et affichage des polygones servant de référence
agri_true=imread('agri_true.tif');
eau_true=imread('eau_true.tif');
foret_true=imread('foret_true.tif');
urbain_true=imread('urbain_true.tif');
sol_true=imread('sol_true.tif');

agri_true(agri_true>0)=1;
agri_ID=find(agri_true==1);
% Les pixels de la classe 1 étant beaucoup plus nombreux, seuls 50'000 sont
% utilisés pour améliorer l'équilibre des classes. Ils sont choisis aléatoirement.
GT_ID=randsample(agri_ID, 50000);
agri_true=zeros(size(agri_true,1),size(agri_true,2));
agri_true(GT_ID)=1;

eau_true(eau_true>0)=2;
eau_true=double(eau_true);
foret_true(foret_true>0)=3;
foret_true=double(foret_true);
sol_true(sol_true>0)=4;
sol_true=double(sol_true);
urbain_true(urbain_true>0)=5;
urbain_true=double(urbain_true);
GT=agri_true+eau_true+foret_true+sol_true+urbain_true;

%% Matrice de confusion
CM = zeros(n,n);
for i = 1:n
    for j=1:n
        CM(i,j) = length(class(class(GT==j)==i));
    end
end

%control 1: la somme des pixels dans la colonne j doit être égale à la somme
%des pixels valant j dans la GT (ground truth). Control1 doit valoir 5x 0
control1=zeros(1,n);
for j=1:n
    control1(j)=sum(CM(:,j))-length(GT(GT==j));
end

%control 2: la somme des pixels dans la ligne i doit être égale à la somme
%des pixels (pour lesquels ont a une GT) classifiés en i. Control2 doit valoir 5x 0
control2=zeros(1,5);
for i=1:n
    control2(i)=sum(CM(i,:))-length(class(class(GT>0)==i));
end

%overall accuracy
OA=sum(diag(CM))/sum(sum(CM));

%user accuracy et producer accuracies
UA=zeros(1,n);
PA=zeros(1,n);
for i=1:n
    UA(i)=CM(i,i)/sum(CM(i,:));
    PA(i)=CM(i,i)/sum(CM(:,i));
end
```

```

%kappa
pobs=sum(diag(CM))/sum(sum(CM));
pest=0;
for i=1:size(CM,1)
    pest=pest+sum(CM(i,:))*sum(CM(:,i))/sum(sum(CM));
end
pest=pest/sum(sum(CM));

kappa=(pobs-pest)/(1-pest);

```

### B.3 Code Matlab utilisé pour le calcul des pourcentages dans les cellules des grilles

```

% Importation de la classification
class=load('class1_matlab.csv');

% Paramètres
res = 10; % résolution spatiale de l'image classifiée en [m]
% coordonnées de la zone d'étude dans le système CH1903
xmin = 484200;
xmax = 516200;
ymin = 107700;
ymax = 136500;

%% ===== Cellules de 3200 m =====
n=9; % nombre de lignes dans la grille
m=10; % nombre de colonnes dans la grille
N=320; %nombre de pixels pour 1 côté de cellule

agri=zeros(n,m);
eau=zeros(n,m);
forets=zeros(n,m);
sol=zeros(n,m);
urbain=zeros(n,m);

%Pour chaque cellule, comptage du nombre de pixels appartenant à chaque classe
for i=1:n
    for j=1:m
        cell=class((i-1)*N+1:i*N,(j-1)*N+1:j*N);
        agri(i,j) = length(cell(cell==1));
        eau(i,j) = length(cell(cell==2));
        forets(i,j)= length(cell(cell==3));
        sol(i,j)= length(cell(cell==4));
        urbain(i,j)=length(cell(cell==5));
    end
end

%Control: dans chaque cellule, la somme des nombres de pixels comptés= 320*320
total1=agri+eau+forets+sol+urbain;
control1 = total1(total1 ~= (N*N))

%Calcul du pourcentage de chaque classe
agri=agri./(N*N);
eau=eau./(N*N);
forets=forets./(N*N);
sol=sol./(N*N);
urbain=urbain./(N*N);

```

```

%Creation d'une table avec tous les résultats
%1. Création d'une matrice avec les coordonnées X et Y du centre de chaque
%cellules, en coordonnées Suisse
X_CH = xmin+N*res/2:N*res:xmax-N*res/2;
X_CH = repmat(X_CH,n,1); %pour avoir 9x X(1) puis 9x X(2),.... X(10)
X_CH = reshape(X_CH,n*m,1);
Y_CH = ymin+N*res/2:N*res:ymax-N*res/2;
Y_CH = sort(Y_CH,'descend'); %pour avoir 10x [Y(9) Y(8) Y(7) Y(6)... Y(1)]
Y_CH = repmat(Y_CH,1,m);
Y_CH = reshape(Y_CH,n*m,1);

%2. Ajout des données calculées
agri=reshape(agri,n*m,1);
eau=reshape(eau,n*m,1);
forets=reshape(forets,n*m,1);
sol=reshape(sol,n*m,1);
urbain=reshape(urbain,n*m,1);

result_3200 = [X_CH Y_CH agri eau forets sol urbain agri+sol agri+forets
agri+sol+forets];
dlmwrite('Class1Matlab_3200.csv',result_3200,'precision','%0.4f')

```

Un code similaire est utilisé pour les autres grilles.

## C Annexe

### C.1 Correspondance entre les classes de la classification et les catégories de la statistique Suisse de la superficie

Les classes "sol" et "végétations basses" ont été regroupées sous le numéro 1. Les points d'interrogation indiquent les catégories qui n'ont pas pu être attribuées.

Classe 1 = végétation basse + Sol

Classe 2 = Eau

Classe 3 = Forêt

Classe 5 = Urbain

Catégorie de la statistique		Classe	Catégorie de la statistique		Classe
1	Bâtiments industriels et artisanaux	5	37	Plantations fruitières	3
2	Terrains attenants aux bâtiments industriels et artisanaux	?	38	Arbres fruitiers de plein champ	3
3	Maisons individuelles et maisons de deux logements	5	39	Vignobles	1
4	Terrains attenants aux maisons individuelles et maisons de deux logements	?	40	Horticulture	1
5	Maisons alignées et en terrasses	5	41	Terres arables	1
6	Terrains attenants aux maisons alignées et en terrasses	?	42	Prairies naturelles	1
7	Immeubles résidentiels	5	43	Pâturages locaux	1
8	Terrains attenants aux immeubles résidentiels	?	44	Prés et pâturages locaux embroussaillés	1
9	Bâtiments publics	5	45	Alpages fauchés	1
10	Terrains attenants aux bâtiments publics	?	46	Alpages favorables	1
11	Bâtiments agricoles	5	47	Alpages embroussaillés	1
12	Terrains attenants aux bâtiments agricoles	?	48	Alpages rocailleux	1
13	Bâtiments non déterminés	5	49	Alpes à moutons	1
14	Terrains attenants aux bâtiments non déterminés	?	50	Forêt normale	3
15	Autoroutes	5	51	Forêt étroite	3
16	Bordures d'autoroutes	?	52	Aires afforestées	?
17	Routes, chemins	?	53	Coupes de bois	?
18	Bordures de routes	?	54	Surfaces forestières dévastées	3
19	Aires de parc de stationnement	5	55	Forêt clairsemée (sur surfaces agricoles)	3
20	Aires ferroviaires stabilisées	5	56	Forêt clairsemée (sur surfaces improductives)	3
21	Bordures de voies ferrées	?	57	Forêt buissonnante	3
22	Aérodromes (surfaces en dur)	5	58	Bosquets, haies	3
23	Aérodromes (surfaces gazonnées)	1	59	Groupes d'arbres (sur surfaces agricoles)	3
24	Installations d'approvisionnement en énergie	5	60	Groupes d'arbres (sur surfaces improductives)	3
25	Stations d'épuration des eaux usées	5	61	Lacs	2
26	Autres installations d'approvisionnement et d'élimination	5	62	Cours d'eau	2
27	Décharges	?	63	Ouvrages de protection contre les crues	?
28	Extraction de matériaux	5	64	Arbustres, broussailles	3
29	Chantiers	5	65	Végétation herbacée improductive	1
30	Friches et bâtiments désaffectés	?	66	Ouvrages de protection contre les chutes de pierres et paravalanches	?
31	Parcs publics	1	67	Zones humides	?
32	Installations de sport	?	68	Infrastructure de sports alpins	5
33	Terrains de golf	1	69	Rochers	5
34	Terrains de camping	?	70	Eboulis, sable	1
35	Jardins familiaux	1	71	Interventions dans le paysage	?
36	Cimetières	?	72	Glaciers, névés	?

## C.2 Correspondance entre les classes de la classification de Florent Cornuet et les catégories de la statistique Suisse de la superficie

Les points d'interrogation indiquent les catégories qui n'ont pas pu être attribuées.

Classe 1 = Eau  
Classe 4 = Végétation

Classe 2 = Surfaces perméables  
Classe 5 = Surfaces imperméables

Classe 3 = Bâtiment

Catégorie de la statistique		Classe	Catégorie de la statistique		Classe
1	Bâtiments industriels et artisanaux	3	37	Plantations fruitières	4
2	Terrains attenants aux bâtiments industriels et artisanaux	?	38	Arbres fruitiers de plein champ	4
3	Maisons individuelles et maisons de deux logements	3	39	Vignobles	2
4	Terrains attenants aux maisons individuelles et maisons de deux logements	?	40	Horticulture	2
5	Maisons alignées et en terrasses	3	41	Terres arables	2
6	Terrains attenants aux maisons alignées et en terrasses	?	42	Prairies naturelles	2
7	Immeubles résidentiels	3	43	Pâturages locaux	2
8	Terrains attenants aux immeubles résidentiels	?	44	Prés et pâturages locaux embroussaillés	2
9	Bâtiments publics	3	45	Alpages fauchés	2
10	Terrains attenants aux bâtiments publics	?	46	Alpages favorables	2
11	Bâtiments agricoles	3	47	Alpages embroussaillés	2
12	Terrains attenants aux bâtiments agricoles	?	48	Alpages rocailleux	2
13	Bâtiments non déterminés	3	49	Alpes à moutons	2
14	Terrains attenants aux bâtiments non déterminés	?	50	Forêt normale	4
15	Autoroutes	5	51	Forêt étroite	4
16	Bordures d'autoroutes	?	52	Aires afforestées	?
17	Routes, chemins	5	53	Coupes de bois	?
18	Bordures de routes	?	54	Surfaces forestières dévastées	4
19	Aires de parc de stationnement	5	55	Forêt clairsemée (sur surfaces agricoles)	4
20	Aires ferroviaires stabilisées	5	56	Forêt clairsemée (sur surfaces improductives)	4
21	Bordures de voies ferrées	?	57	Forêt buissonnante	4
22	Aérodromes (surfaces en dur)	5	58	Bosquets, haies	4
23	Aérodromes (surfaces gazonnées)	2	59	Groupes d'arbres (sur surfaces agricoles)	4
24	Installations d'approvisionnement en énergie	5	60	Groupes d'arbres (sur surfaces improductives)	4
25	Stations d'épuration des eaux usées	5	61	Lacs	1
26	Autres installations d'approvisionnement et d'élimination	5	62	Cours d'eau	1
27	Décharges	?	63	Ouvrages de protection contre les crues	?
28	Extraction de matériaux	?	64	Arbustes, broussailles	2
29	Chantiers	?	65	Végétation herbacée improductive	2
30	Friches et bâtiments désaffectés	?	66	Ouvrages de protection contre les chutes de pierres et paravalanches	?
31	Parcs publics	4	67	Zones humides	?
32	Installations de sport	?	68	Infrastructure de sports alpins	3
33	Terrains de golf	2	69	Rochers	5
34	Terrains de camping	?	70	Eboulis, sable	2
35	Jardins familiaux	2	71	Interventions dans le paysage	?
36	Cimetières	?	72	Glaciers, névés	?

## D Annexe : Résultat des classifications

### D.1 Annexe : Résultat de la classification avec Erdas

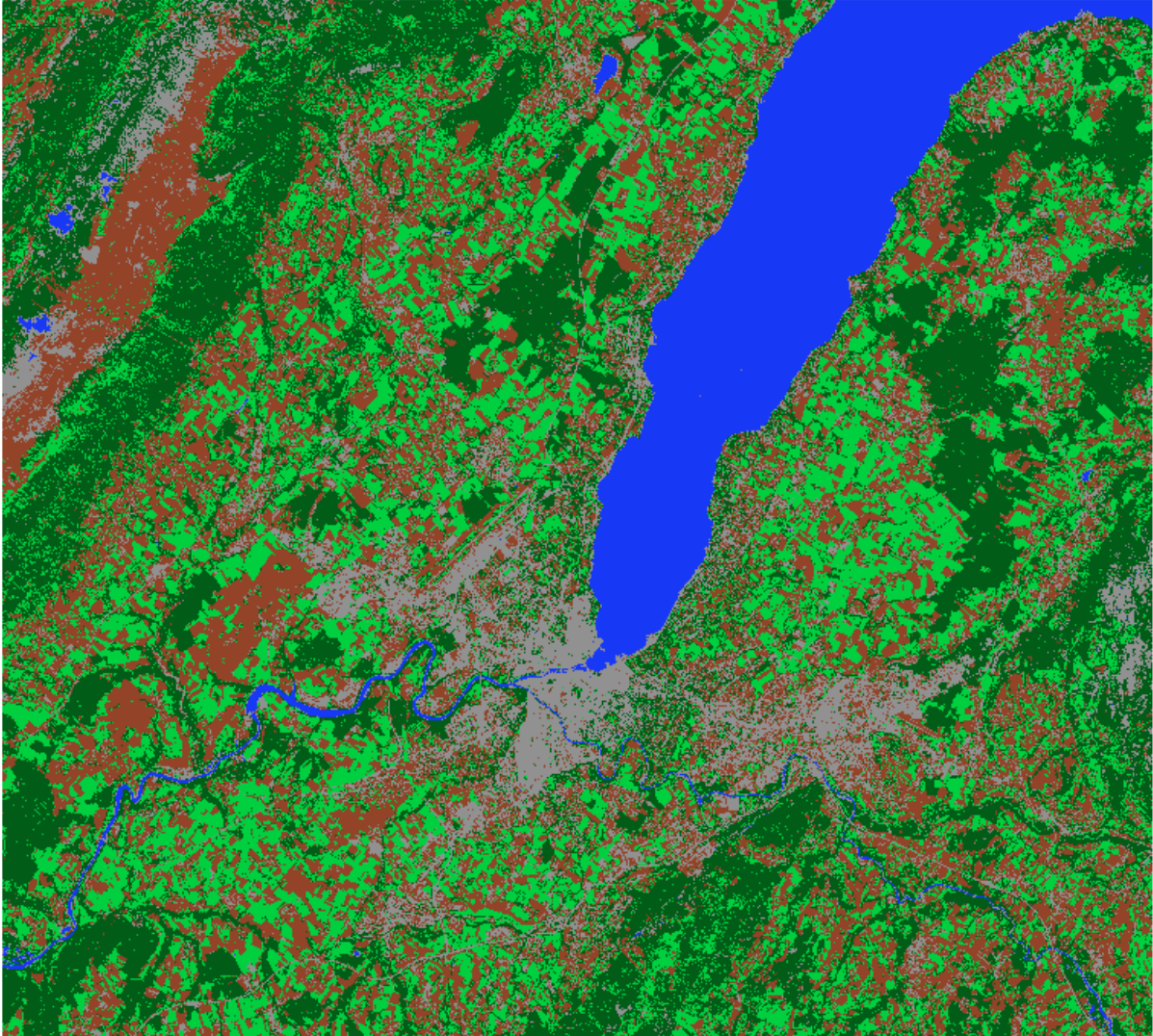
■ Végétation basse   ■ Forêt   ■ Sol   ■ Eau   ■ Urbain





D.2 Annexe : Résultat de la classification avec Matlab

 Végétation basse Forêt Sol Eau Urbain



# E Annexe : Valeurs numériques des corrélations

## E.1 Adultes, Classification réalisée avec Matlab

	vege3200	eau3200	foret3200	sol3200	urbain3200	vs3200	vf3200	vsf3200	vege800	eau800	foret800	sol800	urbain800	vs800	vf800	vsf800	vege200	foret200	eau200	sol200	urbain200	vs200	vf200	vsf200	bmi_corr	bmi	revmed
vege3200	1.000	-0.302	0.138	0.809	-0.762	0.966	0.906	0.928	0.716	-0.185	0.035	0.569	-0.657	0.737	0.582	0.685	0.489	-0.110	0.051	0.565	-0.593	0.659	0.320	0.600	0.018	-0.016	0.413
eau3200	-0.302	1.000	-0.285	-0.597	-0.304	-0.447	-0.378	-0.492	-0.136	0.530	0.039	-0.227	-0.002	-0.212	-0.083	-0.178	-0.087	0.076	0.081	-0.164	0.098	-0.167	0.008	-0.106	-0.016	-0.016	0.000
foret3200	0.138	-0.285	1.000	0.122	-0.209	0.138	0.545	0.411	0.023	-0.130	0.405	0.024	-0.142	0.027	0.259	0.178	0.036	-0.065	0.249	0.017	-0.141	0.030	0.201	0.148	0.006	-0.010	0.206
sol3200	0.809	-0.597	0.122	1.000	-0.475	0.933	0.737	0.894	0.547	-0.345	-0.032	0.663	-0.538	0.701	0.410	0.627	0.366	-0.115	-0.034	0.588	-0.518	0.622	0.188	0.526	0.028	0.003	0.313
urbain3200	-0.762	-0.304	-0.209	-0.475	1.000	-0.674	-0.734	-0.679	-0.575	-0.140	-0.166	-0.459	0.688	-0.593	-0.550	-0.604	-0.400	0.075	-0.162	-0.475	0.556	-0.549	-0.349	-0.558	-0.012	0.025	-0.457
vs3200	0.966	-0.447	0.138	0.933	-0.674	1.000	0.877	0.960	0.677	-0.265	0.007	0.638	-0.638	0.758	0.535	0.694	0.459	-0.118	0.016	0.603	-0.590	0.676	0.278	0.597	0.023	-0.008	0.390
vf3200	0.906	-0.378	0.545	0.737	-0.734	0.877	1.000	0.962	0.616	-0.212	0.203	0.492	-0.617	0.635	0.603	0.656	0.429	-0.121	0.149	0.485	-0.562	0.571	0.357	0.571	0.018	-0.018	0.438
vsf3200	0.928	-0.492	0.411	0.894	-0.679	0.960	0.962	1.000	0.630	-0.281	0.122	0.595	-0.627	0.705	0.566	0.689	0.433	-0.127	0.085	0.560	-0.583	0.631	0.313	0.592	0.023	-0.011	0.418
vege800	0.716	-0.136	0.023	0.547	-0.575	0.677	0.616	0.630	1.000	-0.184	0.037	0.503	-0.770	0.852	0.806	0.791	0.644	-0.106	0.069	0.543	-0.650	0.712	0.425	0.655	0.038	0.002	0.450
eau800	-0.185	0.530	-0.130	-0.345	-0.140	-0.265	-0.212	-0.281	-0.184	1.000	-0.151	-0.318	-0.017	-0.293	-0.233	-0.324	-0.096	0.252	-0.001	-0.201	0.143	-0.199	-0.056	-0.174	-0.017	-0.011	-0.071
foret800	0.035	0.039	0.405	-0.032	-0.166	0.007	0.203	0.122	0.037	-0.151	1.000	0.028	-0.381	0.037	0.622	0.412	0.072	-0.107	0.604	0.007	-0.319	0.038	0.480	0.328	0.012	-0.011	0.286
sol800	0.569	-0.227	0.024	0.663	-0.459	0.638	0.492	0.595	0.503	-0.318	0.028	1.000	-0.746	0.881	0.411	0.814	0.386	-0.165	0.047	0.787	-0.696	0.786	0.258	0.709	0.031	-0.007	0.463
urbain800	-0.657	-0.002	-0.142	-0.538	0.688	-0.638	-0.617	-0.627	-0.770	-0.017	-0.381	-0.746	1.000	-0.873	-0.829	-0.940	-0.560	0.104	-0.305	-0.677	0.825	-0.778	-0.546	-0.828	-0.037	0.011	-0.596
vs800	0.737	-0.212	0.027	0.701	-0.593	0.758	0.635	0.705	0.852	-0.293	0.037	0.881	-0.873	1.000	0.690	0.926	0.587	-0.158	0.066	0.774	-0.778	0.866	0.389	0.788	0.040	-0.003	0.527
vf800	0.582	-0.083	0.259	0.410	-0.550	0.535	0.603	0.566	0.806	-0.233	0.622	0.411	-0.829	0.690	1.000	0.864	0.548	-0.146	0.413	0.430	-0.698	0.581	0.617	0.708	0.037	-0.005	0.522
vsf800	0.685	-0.178	0.178	0.627	-0.604	0.694	0.656	0.689	0.791	-0.324	0.412	0.814	-0.940	0.926	0.864	1.000	0.562	-0.184	0.289	0.709	-0.829	0.804	0.536	0.843	0.041	-0.007	0.588
vege200	0.489	-0.087	0.036	0.366	-0.400	0.459	0.429	0.433	0.644	-0.096	0.072	0.386	-0.560	0.587	0.548	0.562	1.000	-0.079	0.164	0.275	-0.658	0.664	0.700	0.659	0.017	-0.016	0.414
foret200	-0.110	0.076	-0.065	-0.115	0.075	-0.118	-0.121	-0.127	-0.106	0.252	-0.107	-0.165	0.104	-0.158	-0.146	-0.184	-0.079	1.000	-0.084	-0.142	0.036	-0.146	-0.107	-0.168	-0.016	-0.010	-0.070
eau200	0.051	0.081	0.249	-0.034	-0.162	0.016	0.149	0.085	0.069	-0.001	0.604	0.047	-0.305	0.066	0.413	0.289	0.164	-0.084	1.000	-0.095	-0.484	0.000	0.819	0.489	-0.016	-0.038	0.271
sol200	0.565	-0.164	0.017	0.588	-0.475	0.603	0.485	0.560	0.543	-0.201	0.007	0.787	-0.677	0.774	0.430	0.709	0.275	-0.142	-0.095	1.000	-0.732	0.902	0.091	0.740	0.034	-0.007	0.511
urbain200	-0.593	0.098	-0.141	-0.518	0.556	-0.590	-0.562	-0.583	-0.650	0.143	-0.319	-0.696	0.825	-0.778	-0.698	-0.829	-0.658	0.036	-0.484	-0.732	1.000	-0.865	-0.733	-0.991	-0.020	0.032	-0.641
vs200	0.659	-0.167	0.030	0.622	-0.549	0.676	0.571	0.631	0.712	-0.199	0.038	0.786	-0.778	0.866	0.581	0.804	0.664	-0.146	0.000	0.902	-0.865	1.000	0.386	0.872	0.035	-0.013	0.584
vf200	0.320	0.008	0.201	0.188	-0.349	0.278	0.357	0.313	0.425	-0.056	0.480	0.258	-0.546	0.389	0.617	0.536	0.700	-0.107	0.819	0.091	-0.733	0.386	1.000	0.737	-0.002	-0.037	0.436
vsf200	0.600	-0.106	0.148	0.526	-0.558	0.597	0.571	0.592	0.655	-0.174	0.328	0.709	-0.828	0.788	0.708	0.843	0.659	-0.168	0.489	0.740	-0.991	0.872	0.737	1.000	0.022	-0.030	0.642
bmi_corr	0.018	-0.016	0.006	0.028	-0.012	0.023	0.018	0.023	0.038	-0.017	0.012	0.031	-0.037	0.040	0.037	0.041	0.017	-0.016	-0.016	0.034	-0.020	0.035	-0.002	0.022	1.000	0.997	0.015
bmi	-0.016	-0.016	-0.010	0.003	0.025	-0.008	-0.018	-0.011	0.002	-0.011	-0.011	-0.007	0.011	-0.003	-0.005	-0.007	-0.016	-0.010	-0.038	-0.007	0.032	-0.013	-0.037	-0.030	0.997	1.000	-0.066
revmed	0.413	0.000	0.206	0.313	-0.457	0.390	0.438	0.418	0.450	-0.071	0.286	0.463	-0.596	0.527	0.522	0.588	0.414	-0.070	0.271	0.511	-0.641	0.584	0.436	0.642	0.015	-0.066	1.000

## E.2 Adultes, Classification réalisée par Florent Cornuet

	eau800	perm800	bati800	vege800	imp800	vp800	eau200	perm200	bati200	vege200	imp200	vp200	bmi_corr	bmi	revmed
eau800	1.000	-0.349	0.013	-0.513	-0.106	-0.506	0.181	-0.319	0.389	-0.280	0.144	-0.382	-0.039	-0.030	-0.188
perm800	-0.349	1.000	-0.763	0.269	-0.734	0.896	-0.192	0.832	-0.627	0.325	-0.547	0.754	0.070	0.056	0.292
bati800	0.013	-0.763	1.000	-0.476	0.681	-0.809	0.108	-0.631	0.614	-0.387	0.408	-0.656	-0.044	-0.031	-0.278
vege800	-0.513	0.269	-0.476	1.000	-0.340	0.669	-0.137	0.247	-0.476	0.541	-0.254	0.489	0.020	0.007	0.280
imp800	-0.106	-0.734	0.681	-0.340	1.000	-0.723	0.133	-0.604	0.418	-0.349	0.583	-0.615	-0.044	-0.032	-0.256
vp800	-0.506	0.896	-0.809	0.669	-0.723	1.000	-0.212	0.756	-0.704	0.500	-0.539	0.808	0.063	0.046	0.355
eau200	0.181	-0.192	0.108	-0.137	0.133	-0.212	1.000	-0.191	-0.064	-0.137	0.061	-0.210	-0.038	-0.035	-0.063
perm200	-0.319	0.832	-0.631	0.247	-0.604	0.756	-0.191	1.000	-0.699	0.237	-0.570	0.815	0.035	0.019	0.330
bati200	0.389	-0.627	0.614	-0.476	0.418	-0.704	-0.064	-0.699	1.000	-0.611	0.313	-0.835	-0.022	-0.005	-0.356
vege200	-0.280	0.325	-0.387	0.541	-0.349	0.500	-0.137	0.237	-0.611	1.000	-0.596	0.756	0.013	0.001	0.266
imp200	0.144	-0.547	0.408	-0.254	0.583	-0.539	0.061	-0.570	0.313	-0.596	1.000	-0.739	-0.016	-0.004	-0.237
vp200	-0.382	0.754	-0.656	0.489	-0.615	0.808	-0.210	0.815	-0.835	0.756	-0.739	1.000	0.032	0.013	0.381
bmi_corr	-0.039	0.070	-0.044	0.020	-0.044	0.063	-0.038	0.035	-0.022	0.013	-0.016	0.032	1.000	0.999	0.013
bmi	-0.030	0.056	-0.031	0.007	-0.032	0.046	-0.035	0.019	-0.005	0.001	-0.004	0.013	0.999	1.000	-0.035
revmed	-0.188	0.292	-0.278	0.280	-0.256	0.355	-0.063	0.330	-0.356	0.266	-0.237	0.381	0.013	-0.035	1.000

### E.3 Enfants, Classification réalisée avec Matlab

	vege3200	eau3200	foret3200	sol3200	urbain3200	vs3200	vf3200	vsf3200	vege800	eau800	foret800	sol800	urbain800	vs800	vf800	vsf800	vege200	foret200	eau200	sol200	urbain200	vs200	vf200	vsf200	bmi_corr	bmi	revmed
vege3200	1.000	-0.261	0.154	0.809	-0.783	0.966	0.918	0.937	0.701	-0.185	0.002	0.553	-0.637	0.722	0.553	0.665	0.472	-0.110	0.009	0.549	-0.573	0.640	0.294	0.577	0.027	-0.010	0.418
eau3200	-0.261	1.000	-0.206	-0.545	-0.315	-0.400	-0.306	-0.426	-0.130	0.591	0.074	-0.240	-0.040	-0.219	-0.059	-0.174	-0.094	0.039	0.127	-0.174	0.094	-0.177	0.034	-0.097	-0.051	-0.048	-0.024
foret3200	0.154	-0.206	1.000	0.106	-0.257	0.140	0.533	0.395	0.092	-0.088	0.343	0.072	-0.194	0.094	0.273	0.214	0.072	-0.054	0.199	0.098	-0.189	0.109	0.187	0.193	0.055	0.037	0.204
sol3200	0.809	-0.545	0.106	1.000	-0.523	0.934	0.735	0.895	0.524	-0.367	-0.064	0.662	-0.513	0.693	0.375	0.614	0.344	-0.105	-0.061	0.566	-0.498	0.595	0.166	0.503	0.031	0.002	0.329
urbain3200	-0.783	-0.315	-0.257	-0.523	1.000	-0.707	-0.774	-0.724	-0.567	-0.158	-0.126	-0.449	0.682	-0.585	-0.520	-0.585	-0.376	0.096	-0.131	-0.464	0.534	-0.530	-0.323	-0.537	-0.006	0.032	-0.428
vs3200	0.966	-0.400	0.140	0.934	-0.707	1.000	0.883	0.965	0.658	-0.275	-0.027	0.629	-0.614	0.745	0.502	0.676	0.439	-0.114	-0.022	0.584	-0.568	0.652	0.252	0.573	0.031	-0.005	0.400
vf3200	0.918	-0.306	0.533	0.735	-0.774	0.883	1.000	0.960	0.637	-0.194	0.140	0.502	-0.623	0.656	0.584	0.655	0.433	-0.116	0.087	0.510	-0.566	0.592	0.326	0.572	0.046	0.006	0.440
vsf3200	0.937	-0.426	0.395	0.895	-0.724	0.965	0.960	1.000	0.635	-0.278	0.066	0.603	-0.621	0.717	0.538	0.684	0.427	-0.120	0.032	0.568	-0.578	0.634	0.283	0.583	0.043	0.005	0.425
vege800	0.701	-0.130	0.092	0.524	-0.567	0.658	0.637	0.635	1.000	-0.190	0.042	0.484	-0.768	0.841	0.812	0.790	0.613	-0.115	0.051	0.545	-0.648	0.701	0.409	0.652	0.007	-0.036	0.473
eau800	-0.185	0.591	-0.088	-0.367	-0.158	-0.275	-0.194	-0.278	-0.190	1.000	-0.127	-0.349	-0.019	-0.319	-0.224	-0.340	-0.131	0.206	0.001	-0.223	0.186	-0.233	-0.079	-0.208	-0.016	-0.006	-0.116
foret800	0.002	0.074	0.343	-0.064	-0.126	-0.027	0.140	0.066	0.042	-0.127	1.000	0.001	-0.369	0.024	0.618	0.392	0.056	-0.118	0.541	0.016	-0.284	0.038	0.422	0.294	0.012	-0.009	0.228
sol800	0.553	-0.240	0.072	0.662	-0.449	0.629	0.502	0.603	0.484	-0.349	0.001	1.000	-0.729	0.880	0.382	0.810	0.375	-0.187	0.006	0.762	-0.671	0.760	0.232	0.683	0.032	-0.009	0.464
urbain800	-0.637	-0.040	-0.194	-0.513	0.682	-0.614	-0.623	-0.621	-0.768	-0.019	-0.369	-0.729	1.000	-0.867	-0.820	-0.934	-0.524	0.142	-0.244	-0.671	0.791	-0.758	-0.494	-0.796	-0.021	0.030	-0.577
vs800	0.722	-0.219	0.094	0.693	-0.585	0.745	0.656	0.717	0.841	-0.319	0.024	0.880	-0.867	1.000	0.676	0.929	0.564	-0.177	0.031	0.766	-0.766	0.850	0.365	0.775	0.024	-0.025	0.543
vf800	0.553	-0.059	0.273	0.375	-0.520	0.502	0.584	0.538	0.812	-0.224	0.618	0.382	-0.820	0.676	1.000	0.851	0.515	-0.159	0.356	0.439	-0.676	0.574	0.569	0.685	0.012	-0.034	0.505
vsf800	0.665	-0.174	0.214	0.614	-0.585	0.676	0.655	0.684	0.790	-0.340	0.392	0.810	-0.934	0.929	0.851	1.000	0.540	-0.207	0.229	0.711	-0.810	0.796	0.493	0.822	0.026	-0.026	0.584
vege200	0.472	-0.094	0.072	0.344	-0.376	0.439	0.433	0.427	0.613	-0.131	0.056	0.375	-0.524	0.564	0.515	0.540	1.000	-0.090	0.131	0.274	-0.660	0.668	0.703	0.661	0.013	-0.021	0.377
foret200	-0.110	0.039	-0.054	-0.105	0.096	-0.114	-0.116	-0.120	-0.115	0.206	-0.118	-0.187	0.142	-0.177	-0.159	-0.207	-0.090	1.000	-0.079	-0.161	0.069	-0.166	-0.111	-0.186	-0.027	-0.017	-0.106
eau200	0.009	0.127	0.199	-0.061	-0.131	-0.022	0.087	0.032	0.051	0.001	0.541	0.006	-0.244	0.031	0.356	0.229	0.131	-0.079	1.000	-0.125	-0.445	-0.037	0.798	0.448	-0.010	-0.026	0.182
sol200	0.549	-0.174	0.098	0.566	-0.464	0.584	0.510	0.568	0.545	-0.223	0.016	0.762	-0.671	0.766	0.439	0.711	0.274	-0.161	-0.125	1.000	-0.736	0.899	0.077	0.744	0.021	-0.028	0.549
urbain200	-0.573	0.094	-0.189	-0.498	0.534	-0.568	-0.566	-0.578	-0.648	0.186	-0.284	-0.671	0.791	-0.766	-0.676	-0.810	-0.660	0.069	-0.445	-0.736	1.000	-0.871	-0.721	-0.993	-0.012	0.043	-0.618
vs200	0.640	-0.177	0.109	0.595	-0.530	0.652	0.592	0.634	0.701	-0.233	0.038	0.760	-0.758	0.850	0.574	0.796	0.668	-0.166	-0.037	0.899	-0.871	1.000	0.380	0.877	0.023	-0.031	0.597
vf200	0.294	0.034	0.187	0.166	-0.323	0.252	0.326	0.283	0.409	-0.079	0.422	0.232	-0.494	0.365	0.569	0.493	0.703	-0.111	0.798	0.077	-0.721	0.380	1.000	0.723	0.001	-0.031	0.360
vsf200	0.577	-0.097	0.193	0.503	-0.537	0.573	0.572	0.583	0.652	-0.208	0.294	0.683	-0.796	0.775	0.685	0.822	0.661	-0.186	0.448	0.744	-0.993	0.877	0.723	1.000	0.015	-0.040	0.621
bmi_corr	0.027	-0.051	0.055	0.031	-0.006	0.031	0.046	0.043	0.007	-0.016	0.012	0.032	-0.021	0.024	0.012	0.026	0.013	-0.027	-0.010	0.021	-0.012	0.023	0.001	0.015	1.000	0.996	0.005
bmi	-0.010	-0.048	0.037	0.002	0.032	-0.005	0.006	0.005	-0.036	-0.006	-0.009	-0.009	0.030	-0.025	-0.034	-0.026	-0.021	-0.017	-0.026	-0.028	0.043	-0.031	-0.031	-0.040	0.996	1.000	-0.084
revmed	0.418	-0.024	0.204	0.329	-0.428	0.400	0.440	0.425	0.473	-0.116	0.228	0.464	-0.577	0.543	0.505	0.584	0.377	-0.106	0.182	0.549	-0.618	0.597	0.360	0.621	0.005	-0.084	1.000

#### E.4 Enfants, Classification réalisée par Florent Cornuet

	eau800	perm800	bati800	vege800	imp800	vp800	eau200	perm200	bati200	vege200	imp200	vp200	bmi_corr	bmi	revmed
eau800	1	-0.3584	0.02776	-0.4849	-0.1264	-0.5134	0.16138	-0.3309	0.44379	-0.302	0.14887	-0.4171	-0.0907	-0.0832	-0.1432
perm800	-0.358	1.000	-0.760	0.176	-0.697	0.892	-0.237	0.826	-0.604	0.216	-0.511	0.735	0.036	0.021	0.272
bati800	0.028	-0.760	1.000	-0.405	0.640	-0.802	0.179	-0.621	0.581	-0.266	0.330	-0.612	0.009	0.025	-0.289
vege800	-0.485	0.176	-0.405	1.000	-0.289	0.601	-0.155	0.162	-0.416	0.508	-0.172	0.409	0.057	0.043	0.259
imp800	-0.126	-0.697	0.640	-0.289	1.000	-0.698	0.188	-0.558	0.320	-0.240	0.550	-0.550	-0.021	-0.007	-0.258
vp800	-0.513	0.892	-0.802	0.601	-0.698	1.000	-0.264	0.745	-0.681	0.408	-0.494	0.784	0.055	0.037	0.340
eau200	0.161	-0.237	0.179	-0.155	0.188	-0.264	1.000	-0.216	-0.010	-0.104	0.063	-0.220	-0.033	-0.029	-0.070
perm200	-0.331	0.826	-0.621	0.162	-0.558	0.745	-0.216	1.000	-0.701	0.151	-0.566	0.827	0.021	0.005	0.291
bati200	0.444	-0.604	0.581	-0.416	0.320	-0.681	-0.010	-0.701	1.000	-0.542	0.259	-0.827	-0.040	-0.022	-0.335
vege200	-0.302	0.216	-0.266	0.508	-0.240	0.408	-0.104	0.151	-0.542	1.000	-0.535	0.680	0.021	0.009	0.235
imp200	0.149	-0.511	0.330	-0.172	0.550	-0.494	0.063	-0.566	0.259	-0.535	1.000	-0.724	0.012	0.022	-0.197
vp200	-0.417	0.735	-0.612	0.409	-0.550	0.784	-0.220	0.827	-0.827	0.680	-0.724	1.000	0.028	0.009	0.349
bmi_corr	-0.091	0.036	0.009	0.057	-0.021	0.055	-0.033	0.021	-0.040	0.021	0.012	0.028	1.000	0.999	0.033
bmi	-0.083	0.021	0.025	0.043	-0.007	0.037	-0.029	0.005	-0.022	0.009	0.022	0.009	0.999	1.000	-0.021
revmed	-0.143	0.272	-0.289	0.259	-0.258	0.340	-0.070	0.291	-0.335	0.235	-0.197	0.349	0.033	-0.021	1.000