

Remerciements

Je tiens tout d'abord à remercier mon maître de stage, Stéphane Joost, pour m'avoir accueilli pendant 5 mois au sein du LASIG et pour sa disponibilité qui a permis le bon déroulement de ce stage.

Je remercie également Claudio Carneiro pour ses conseils précieux et son aide sur le LiDAR.

Mes remerciements vont également à Alain Clappier pour m'avoir transmis son intérêt pour les modèles atmosphériques.

Je tiens à remercier Nadège Blond pour avoir organisé mon séjour à Strasbourg.

Je tiens tout particulièrement à remercier Dasaraden Mauree pour sa disponibilité et sa patience pour m'avoir expliquer le fonctionnement du modèle météorologique et permis d'obtenir les résultats produits par le modèle.

Mes remerciements s'adressent également à mon Mme Élisabeth Simonetto pour les améliorations qu'elle a pu apporter à ce document.

SOMMAIRE

1.Introduction.....	2
2.État de l'art.....	5
<i>a. Modèle FVM.....</i>	<i>5</i>
<i>b. LiDAR.....</i>	<i>10</i>
<i>c. LANDSAT dans le cadre de l'utilisation des modèles FVM....</i>	<i>18</i>
3.Méthodes utilisées.....	18
<i>a. Méthode de classification des images LANDSAT.....</i>	<i>19</i>
<i>b. Méthode de classification des données LiDAR.....</i>	<i>26</i>
<i>c. Méthode de calcul des données d'utilisation du sol.....</i>	<i>38</i>
4.Résultats.....	40
a. Résultats LiDAR.....	40
b. Résultats LANDSAT.....	42
c. Comparaison LiDAR-LANDSAT.....	44
5.Discussion.....	46
6.Conclusion.....	49
Références bibliographiques.....	50
Annexes.....	52

Introduction

Avec une forte urbanisation et une croissance des activités humaines, la pollution de l'air est un thème qui concerne de plus en plus les sociétés actuelles. L'émission des polluants dans l'atmosphère ne cesse d'augmenter entraînant des problèmes sanitaires et environnementaux (maladies respiratoires et cardiovasculaires, cancers, réchauffement climatique...). Les polluants sont issus essentiellement des véhicules à moteurs, des installations industrielles, des centrales thermiques, des activités agricoles et des installations de chauffages et de climatisation. Une grande partie de cette pollution est concentrée dans les villes et elle a un impact important sur l'atmosphère. Pour répondre à cette problématique, des modèles numériques sont créés pour améliorer la compréhension, pour anticiper la pollution en milieu urbain ainsi que dans le but de proposer des solutions visant à réduire ces impacts (dégradation de la qualité de l'air, hausse des températures).

Comme la diffusion des polluants dépend fortement des conditions météorologiques, une première étape de la modélisation de la pollution de l'air s'appuie sur des modèles météorologiques.

Un de ces modèles a été développé à l'EPFL (École Polytechnique Fédérale de Lausanne). Sa particularité est de tenir compte de l'influence des surfaces urbaines. C'est un modèle méso-échelle FVM (Finite Volume Model) qui calcule les champs météorologiques (pression, température, humidité, vitesse et orientation du vent) dans chacune des mailles du modèle en tenant compte des effets de la ville sur l'atmosphère (MARTILLI, 2001). L'objet de cette présente étude est de le faire fonctionner en lui permettant d'insérer au bon format les données d'utilisation du sol.

En général, ces données d'utilisation du sol sont issues d'une classification basée sur des images LANDSAT ayant une résolution de 30m. Une grille régulière est superposée à cette classification et les pourcentages surfaciques de chaque classe sont reportés dans chacune des cellules pour constituer les données d'utilisation du sol prises en compte par le modèle.

La présente recherche crée des données d'utilisation du sol à partir du LiDAR aéroporté pour estimer les apports potentiels de ce type de données pour les modèles de pollution de l'air décrits ci-dessus. Le LiDAR permet principalement d'apporter des informations liées à l'altimétrie. Il présente également une meilleure résolution spatiale (1m dans notre cas) et une meilleure précision (15 cm à 20cm en altimétrie et 30 cm à 40cm en planimétrie). Ainsi, son utilisation pourrait permettre d'affiner les prédictions des modèles méso-échelle.

La pollution de l'air est un problème important et toujours d'actualité dans le canton de Genève (plusieurs fois par an, la pollution dépasse les normes autorisées), raison pour laquelle nous avons décidé de classifier le mode d'utilisation du sol de cette ville afin de nourrir le modèle de diffusion des polluants correspondants. Deux modèles d'utilisation du sol seront calculés, l'un sur la base de la données LiDAR combinée avec une carte 2D du bâti, et l'autre sur la base d'images LANDSAT 7.

Au final, les résultats fournis par les deux modèles seront comparés pour déterminer si le LiDAR permet d'améliorer ou non les modèles atmosphériques FVM à méso-échelle de pollution de l'air.

La première partie du mémoire passe en revue les états de l'art sur les modèles FVM, en LiDAR et pour LANDSAT dans le cadre de l'utilisation des modèles FVM.

La deuxième partie expose la méthode qui a permis de produire la classification s'appuyant sur les images LANDSAT, la méthode pour créer la classification qui se base sur les données LiDAR combinées avec l'utilisation d'une carte 2D du bâti, et la méthode pour calculer les données d'utilisation du sol pour qu'elles puissent être utilisables par le modèle FVM.

La troisième partie présente les résultats issus des données LiDAR et LANDSAT ainsi que la comparaison de ces résultats fournis par le modèle FVM.

La quatrième partie propose une discussion sur les résultats obtenus.

Dans la dernière partie, nous concluons sur la contribution du LiDAR avec l'utilisation d'une carte 2D du bâti pour une telle application.

État de l'art

Cette partie a pour objet de passer en revue un état des connaissances existantes sur les sujets d'étude traités dans ce mémoire. Pour commencer, un état de l'art sur le modèle FVM est présenté pour comprendre son fonctionnement. Un second état de l'art sur le LiDAR expose l'emploi de cette technique pour produire des classifications des modes d'occupation du sol ou des modes d'utilisation du sol. Un dernier état de l'art présente l'utilisation d'images LANDSAT dans le cadre du modèle FVM.

a. Modèle FVM

Pour pouvoir comprendre comment fonctionne le modèle, il est intéressant d'expliquer les mécanismes qui régissent la dispersion des polluants dans l'air et comment le modèle est conçu pour en tenir compte.

C'est pourquoi, avant de le présenter, nous allons expliquer les mécanismes qui ont des impacts sur la pollution de l'air. Celle-ci dépend de quatre facteurs : les caractéristiques des surfaces urbaines, les conditions météorologiques, les sources de pollution et la topographie.

Les surfaces urbaines produisent des effets mécaniques et des effets thermiques (Figure1) sur l'atmosphère.

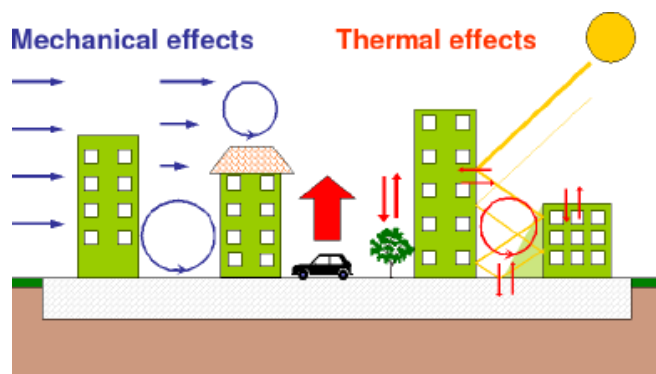


Figure 1: effets mécaniques et thermiques produits par les villes (KRPO, 2009)

L'effet mécanique résulte de la résistance des bâtiments et du sol sur le déplacement des masses d'air. La canopée urbaine a pour conséquence de modifier la vitesse du vent et de former des tourbillons (cf. Figure1 à gauche).

Cet effet mécanique dépend de nombreux paramètres tels que les caractéristiques des rues et des bâtiments (hauteurs, dimensions, formes, couleurs et types de matériaux, orientations). À l'échelle de la ville, la rugosité des bâtiments et des sols contribue à diminuer la vitesse du vent, ce qui réduit le brassage de l'air et contribue à augmenter la pollution de l'air.

L'effet thermique résulte de nombreux paramètres. Les radiations solaires qui

sont rétro-diffusées par les bâtiments ont tendance à bloquer la chaleur (cf. Figure 1 à droite). L'urbanisation modifie l'albédo. Le chauffage ou la climatisation des bâtiments, la combustion de fuel des véhicules ou des installations industrielles sont autant de dégagements de chaleur. La végétation qui a un effet thermo-régulateur rafraîchit l'atmosphère en créant de l'ombrage et en provoquant de l'évapotranspiration qui va transférer une quantité d'eau du sol vers l'atmosphère (KRPO, 2009). Son absence en ville réduit cette action et contribue à un réchauffement du climat urbain.

Les effets thermiques et mécaniques sont responsables de l'apparition de l'îlot de chaleur urbain (cf. Figure 2). C'est un flux de chaleur qui augmente la température de l'air de 0° à 7° au-dessus des villes en fonction de la morphologie de la ville, de la saison et des radiations solaires (Chow and Roth 2006).

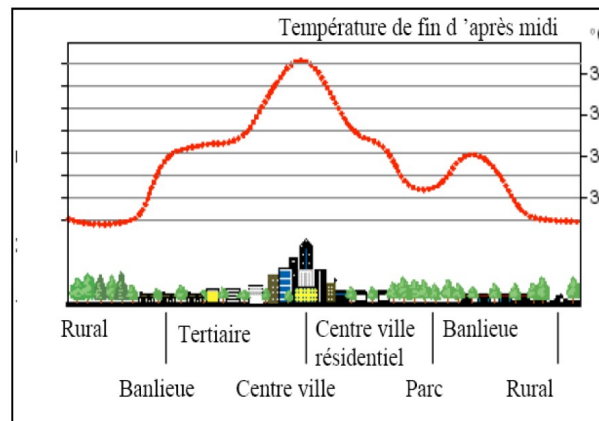


Figure 2: profil schématique de l'îlot de chaleur urbain (source : Heat Island Group)

Cet îlot de chaleur urbain contribue à l'apparition du « smog » qui est un nuage brunâtre composé d'un mélange de polluants présents en période de forte chaleur. Cet îlot pose un problème sanitaire en été. Il conduit à une augmentation du taux de mortalité (maladie cardio-vasculaire et respiratoire, hyperthermie, déshydratation...) lors des canicules car il renforce et allonge les vagues de chaleur et d'humidité. Une solution existe en matière d'urbanisme pour limiter ce phénomène. Elle consiste à privilégier une ouverture des espaces pour éviter le piégeage des radiations dans les canyons urbains et à faciliter le déplacement des masses d'airs. De plus, limiter l'étalement urbain permet de conserver une quantité de végétation importante qui va réguler la température de l'atmosphère.

La dispersion et la concentration des polluants sont directement liées aux conditions météorologiques qui règnent dans le milieu considéré. La concentration de polluants est inversement proportionnelle à la vitesse du vent (source OMS, 2000).

En cas de vents forts, le brassage de l'air crée des conditions climatiques homogènes sur l'ensemble de la ville. La température, les conditions de vents et les taux d'humidité sont globalement identiques. Ce brassage d'air permet de disperser et diluer les polluants.

En cas de ciel couvert, la canopée urbaine est exposée à moins de rayonnement solaire, ce qui a également tendance à limiter l'apparition de micro-climats à l'intérieur de la ville.

Par contre, en cas de vents faibles et de ciel clair, le brassage de l'air est

limité. Ainsi, la dispersion et la dilution des polluants diminuent. C'est dans cette situation que les pics de pollution sont les plus intenses. Des micro-climats se distinguent à l'intérieur de la ville selon les modes d'utilisation du sol (DUPONT, 2001). Dans des zones en présence de végétation, le micro-climat sera plus frais que dans des zones fortement urbanisées où les effets mécaniques et thermiques des surfaces urbaines seront accentués. Cette différence de température a pour conséquence de créer des mouvements de masses d'air depuis les surfaces urbaines qui ont un micro-climat chaud vers celles qui ont un micro-climat plus froid. Ce phénomène crée un courant d'air artificiel de l'air froid vers l'air chaud, de même origine que celui de la brise présente en bord de mer ou sur le flanc des vallées.

Pendant des périodes pluvieuses, les polluants présents dans l'atmosphère sont ramenés au sol. Une partie de ces polluants repartiront dans l'atmosphère une fois déshydratés.

Les émissions de polluants dans l'atmosphère sont corrélées avec l'intensité des activités humaines qui se concentrent en grande majorité dans les zones urbaines.

On peut distinguer deux types de polluants :

- les oxydes d'azote (NO_x)
- les Composés Organiques Volatiles (hydrocarbures)

Dans un premier temps, les oxydes d'azote sont libérés dans l'atmosphère par l'intermédiaire du trafic routier et du chauffage tandis que les composés organiques volatiles émanent des activités industrielles (ROULET, 2004).

Dans un second temps, en présence de rayonnement solaire, ces polluants créent une réaction chimique qui est à l'origine de l'ozone troposphérique (O_3). C'est ce polluant secondaire qui crée le plus d'effets néfastes sur la santé humaine, l'environnement et les infrastructures.

Le relief a des effets locaux sur les conditions de vents et donc sur la diffusion des polluants.

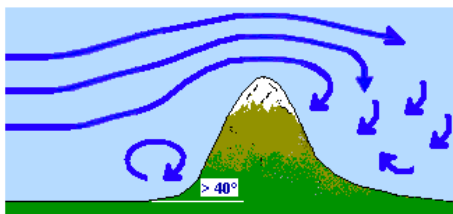


Figure 3: influence du relief en présence de vents synoptiques (LUXORION, 2011)

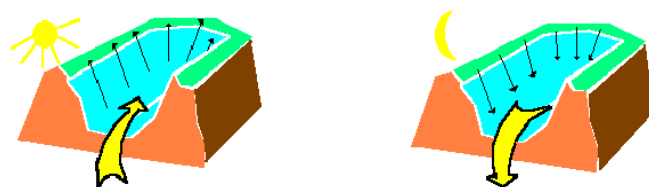


Figure 4: brise de pente (flèche noire) et brise de vallée (flèche jaune) pendant le jour et la nuit en l'absence de vents synoptiques (LUXORION, 2011)

Le relief constitue un obstacle qui entraîne un changement de direction, de vitesse du vent synoptique (vent de hautes altitudes) et l'apparition de turbulences (cf. Figure 3).

Les pentes créent un phénomène de brise (cf. Figure 4). Pendant la journée, le sol, le long des pentes, est réchauffé par le rayonnement solaire qui à son tour va chauffer la couche d'air environnante. Cette couche d'air va avoir ainsi tendance à monter le long de la pente. Pendant la nuit, c'est l'inverse qui va se produire. Le sol va refroidir la couche d'air environnante. Cette couche d'air descend ainsi le long de la pente.

Dans les vallées se produit également un phénomène de brise (cf. Figure 4). Sur chacun des flancs de la vallée se produisent des brises de pentes qui créent un

appel d'air à l'intérieur même de la vallée et amènent de l'air chaud pendant la journée au fond de la vallée. Pendant la nuit, le mouvement d'air va s'inverser.

Pour tenir compte des facteurs présentés ci-dessus, la modélisation de la pollution de l'air se base sur l'utilisation de deux modèles.

Le premier est le modèle météorologique qui simule les conditions météorologiques lors de l'épisode de pollution à étudier. L'objectif est de calculer dans chaque maille du modèle les champs météorologiques (pression, température, humidité, vitesse et orientation du vent...).

Il existe de nombreux modèles de flux atmosphériques simulant des phénomènes météorologiques à différentes échelles. Le modèle utilisé pour ce TFE est le modèle « Finite Volume Model » (Clappier et al. 1996). C'est un modèle méso-échelle, c'est-à-dire un modèle qui décrit des phénomènes qui ont une échelle de temps et d'espace plus grande que des nuages traités individuellement et plus petite que la circulation atmosphérique générale. L'échelle horizontale correspond à une zone de l'ordre de quelques kilomètres jusqu'à plusieurs centaines de kilomètres. L'échelle verticale s'étend de quelques dizaines de mètres jusqu'à l'épaisseur de la troposphère (12km). L'échelle de temps est comprise entre 1 heure et 72 heures.

Ce modèle tridimensionnel se base sur la discrétisation des espaces par la méthode des volumes finis. Cette méthode consiste à appliquer un maillage dans l'espace. Il permet ainsi de résoudre des équations à dérivées partielles complexes basées sur les principes de conservation de masse, de quantité de mouvement, d'énergie et d'humidité de l'air. Pour faire ces calculs, le modèle utilise comme paramètres les données globales météorologiques (appelées "forcig" dans le modèle), les données topographiques locales ("topo"), les données d'utilisation du sol ("landuse") ainsi que des caractéristiques des bâtiments (cf figure5).

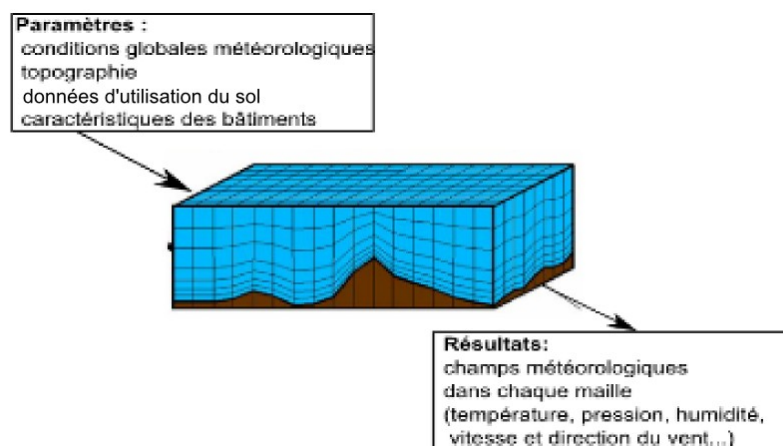


Figure 5: paramètres et résultats du modèle FVM

La particularité de ce modèle FVM est de prendre en compte les effets des surfaces urbaines sur l'atmosphère (Martilli, 2001), du phénomène d'îlot de chaleur urbain de manière plus précise que les méthodes généralement employées (RAMS, Pielk et al 1992 ; MM5, Grell et al, 1994). Pour cela, le modèle fait appel à deux types de paramétrisations : une paramétrisation rurale et une autre urbaine. Les données d'utilisation du sol vont permettre de distinguer les zones rurales des zones urbaines pour déterminer comment les paramétrisations doit s'appliquer dans chacune des mailles du modèle. Il tient compte également de la chaleur émise par les bâtiments et des échanges de chaleur qui se produisent entre les bâtiments et

l'atmosphère (Krpó 2009). Ainsi, le modèle FVM dispose d'un module d'énergie des bâtiments couplé à la paramétrisation des canopées urbaines.

Le modèle FVM se décompose en 3 répertoires : répertoire « input », répertoire « programme » et répertoire « studies ». Le premier répertoire constitue la base de donnée utilisée par le modèle et dans lequel se trouve le modèle numérique de terrain et les données météorologiques au niveau mondial ainsi que les données d'utilisation du sol. Le répertoire « programme » est composé du programme FVM en lui même avec les modules codés en langage FORTRAN. Il y a également dans ce répertoire les préprocesseurs (« preland », « pretopo » et « premeteo ») qui permettent d'aller chercher les données dans le répertoire « INPUT » et de les transformer pour que ces données puissent être utilisables par le modèle. Le dernier répertoire « studies » stocke les résultats et permet d'accéder à un fichier de paramètres (.par). C'est dans ce fichier de paramètre que l'utilisateur va définir la durée et le pas de la simulation, les coordonnées de la zone à étudier et le pas du maillage. Une compilation va être réalisée pour définir le domaine de la simulation (pas du maillage, nombre de maille au niveau horizontal et vertical) qui va être introduits dans les modules du programme. Ensuite, les préprocesseurs (« preland », « pretopo » et « premeteo ») sont lancés pour faire le lien entre le fichier de paramètre (.par) et le fichier exécutable du modèle FVM. Puis le modèle est lancé et les résultats peuvent ainsi être visualisé à l'aide du logiciel FERRET qui a été produit par la NOAA (agence météorologique et océanographique des États-Unis).

Le modèle demande comme paramètre d'entrée les données globales météorologiques, les données topographiques qui sont issues d'un modèle numérique de terrain et les données d'utilisation du sol. Comme dans ce TFE, un des objectifs est de produire des données d'utilisation du sol pour le modèle FVM, le paragraphe suivant explique plus particulièrement la production des données d'utilisation du sol et son traitement par le modèle FVM.

Les données d'utilisation du sol sont obtenues en deux étapes. Une première étape consiste à réaliser la classification des différents modes d'utilisation du sol. Une deuxième étape consiste à superposer une grille sur cette classification puis de calculer dans chaque cellule (1km*1km en général) de la grille le pourcentage surfacique de chaque classe. À l'issue de cette étape, les données d'utilisation du sol constituent un fichier texte qui contient les coordonnées de la cellule et le pourcentage surfacique pour chacune des classes de la cellule.

Ensuite un fichier de paramètres des classes urbaines va permettre de définir les caractéristiques de ces classes. En effet, le principe de la paramétrisation urbaine du modèle consiste à appliquer dans chaque cellule un canyon urbain constitué de deux rues perpendiculaires avec des bâtiments de part et d'autres. La configuration de ce canyon urbain dépend de la classe d'utilisation associée à la cellule. C'est pourquoi, un fichier de paramètres (.par) présent dans le répertoire « programme » est préalablement modifié pour définir la configuration du canyon urbain pour chaque classe urbaine. Ce fichier permet de décrire la configuration des rues (longueur, largeur, orientation, chaleur spécifique et conductivité thermique des matériaux), la configuration des bâtiments (hauteur et leur distribution, largeur, chaleur spécifique et conductivité thermique, albédo, émissivité du toit, des murs, du sol ainsi que la caractéristiques de chauffage et de ventilation des bâtiments). Les caractéristiques de chacune des surfaces (bâtiments, rues) sont déterminés à partir de données issues de la littérature (MARTILLI, 2002). Après avoir configuré le fichier de paramètre des classes urbaines, le « landuse » est prêt pour être inséré dans le modèle. Le modèle FVM va réaliser un calcul de flux pour chaque classe et les

pourcentages surfaciques présents dans les données d'utilisation du sol vont permettre de pondérer ces calculs pour tenir compte de l'influence des différents modes d'utilisation du sol présents dans les cellules.

Des travaux de recherche sont actuellement en cours pour développer le modèle FVM. L'idée est de mettre en place un système qui permette d'intégrer des données issues d'un modèle micro-échelle à l'intérieur du modèle méso-échelle. Ce couplage d'un modèle de bâtiment avec le modèle météorologique permettra de simuler des conditions météorologiques à l'intérieur d'un modèle méso-échelle en affinant son maillage dans certains quartiers de la ville à l'aide du modèle micro-échelle.

Une fois que les champs météorologiques sont calculés dans chacune des mailles du modèle météorologique FVM, un second modèle utilise ces champs météorologiques comme paramètre d'entrée pour simuler la qualité de l'air. Au final, ce modèle photochimique va fournir la concentration des polluants dans chacune des mailles du modèle. Le second modèle fréquemment utilisé est le modèle CHIMERE (Chemistry transport model), car ce modèle de calcul de pollution de l'air est également un modèle méso-échelle.

b. LiDAR

Un objectif de cette étude était de réaliser une classification en s'appuyant sur des données LiDAR. C'est pourquoi, cette partie a pour objet de présenter la technologie LiDAR et son utilisation pour produire une classification du mode d'occupation du sol ou du mode d'utilisation du sol. L'occupation du sol correspond à la couverture de la surface terrestre et qui est directement observable sans interprétation de l'utilisateur. L'occupation du sol comprend par exemple, l'eau, la végétation, les forêts de feuillus, les forêts de conifères, le sable, le sol nu, la prairie, les routes, le bâti dense, le bâti lâche... Les classifications d'occupation du sol sont couramment appelées en anglais des « landcover ». Contrairement à l'occupation du sol, l'utilisation du sol donne une description des fonctions que l'homme exerce sur un espace considéré. Il demande une interprétation de l'utilisateur et est souvent moins détaillé. L'utilisation du sol peut être une zone agricole, urbaine, industrielle... Les classifications d'utilisation du sol en anglais sont désignés comme des « landuse ».

Description du système LiDAR aéroporté

Le LiDAR (Light Detecting And Ranging) est une technologie basée sur l'utilisation d'un capteur actif qui mesure et analyse les propriétés d'un rayonnement lumineux émis par un laser et réfléchi par la surface terrestre. Lorsque le capteur est embarqué dans un avion, un hélicoptère ou un drone, on parle de LiDAR aéroporté.

Le LiDAR est une technique de positionnement des éléments qui composent la surface terrestre. Pour une telle application, on parle de LiDAR topographique. Cette technique produit un nuage de points 3D (X, Y, Z) avec des attributs supplémentaires comme, par exemple, l'intensité et l'écho de chaque signal émis

par le capteur actif. Le LiDAR utilise une source laser qui émet un rayonnement lumineux dans le domaine du visible ou du proche infrarouge contrairement au RaDAR (Radio Detecting and Ranging) qui utilise des ondes radios. Ce rayonnement lumineux est ensuite rétrodiffusé par un objet ou une surface qui est renvoyé en direction du capteur. Ce capteur mesure le temps de propagation aller-retour du rayonnement lumineux et permet de déterminer la distance entre la cible et l'émetteur. Couplée avec un système de positionnement GPS et un système de navigation inertiel INS (cf. Figure 6), on peut déterminer la position du capteur et la direction des impulsions lasers, et ainsi calculer la position d'éléments composant la surface terrestre. La mesure des points lasers ainsi obtenus a une précision d'environ 15cm en altimétrie et 20 à 40cm en planimétrie (Baltsavias, 1999).

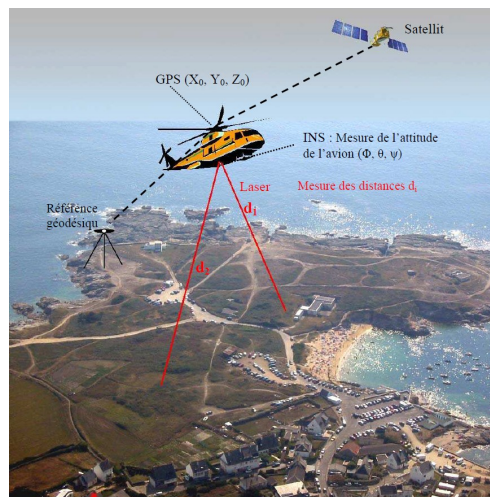


Figure 6: composants du système LiDAR aéroporté (source : TARSHA KURDI, 2008)

Il existe deux techniques de LiDAR topographiques. La première est le système multi-écho. Son principe est de détecter le maxima de l'intensité sur un signal. Ce maxima d'amplitude est appelé écho. Sa détection suffit pour déterminer la distance entre un objet et le capteur pour chacun des signaux et ainsi produire un nuage de points en 3D.

Un autre système LiDAR topographique appelé « à retour d'onde complète » a été développé ces dernières années. Au lieu de mesurer uniquement le maxima des intensités comme c'est le cas pour le système multi écho, la totalité du signal rétrodiffusé est enregistrée. Ainsi, la scène n'est plus constituée d'un nuage de points 3D mais plutôt d'une succession de signaux à une dimension. Ces signaux vont être décomposés en somme de fonctions gaussiennes généralisées (MALLET, 2010).

Une fonction gaussienne généralisée est de la forme $f_{CG} = A * e^{-\frac{(x-\mu)^\alpha}{2\omega^2}}$ avec A :

l'amplitude, ω : la largeur du modèle gaussien, μ indique la position du mode de la fonction gaussienne et α caractérise la forme de cette fonction. Cette décomposition permet de détecter un écho sur le signal avec le paramètre μ et de caractériser la forme de l'écho avec les paramètres A, ω , α . Avec cette technique, il est alors possible de générer un nuage de points 3D, puisque l'on a la distance entre l'objet et la cible avec le paramètre μ , et chacun de ces points LiDAR contient, en plus, des attributs supplémentaires liés à la forme de l'écho (A, ω , α). Il a été montré que ce système permet de densifier le nuage de point 3D en détectant un plus grand nombre d'échos lors de la décomposition du signal et qu'il permet de classifier de

manière plus précise les différentes espèces végétales, les matériaux des toits et d'effectuer une distinction entre le sol naturel et artificiel (MALLET, 2010).

La particularité du LiDAR est de pouvoir enregistrer plusieurs rétrodiffusions avec une seule impulsion suivant la géométrie de la surface. C'est le cas de la végétation où plusieurs échos vont être créés. Une partie de l'impulsion sera rétrodiffusée par le feuillage puis une autre partie de l'impulsion va traverser le feuillage et va être réfléchi par le sol. Le LiDAR est donc particulièrement intéressant pour produire un MNT (Modèle Numérique de Terrain) en particulier dans des zones boisées où la photogrammétrie ne permet pas de visualiser le terrain.

Dans le cas la reconstruction des villes en 3D, le LiDAR est utilisé pour représenter le milieu urbain sous forme d'image où chaque pixel correspond à une élévation. On parle de représentation en 2,5 dimensions. Il est également utilisé pour les représenter sous forme de « modèle bloc ». C'est un modèle en 3 dimensions où les bâtiments sont représentés sous forme de blocs sans structure de toits. L'intérêt du LiDAR aéroporté dans la reconstruction des villes en 3D est de couvrir une zone au niveau urbain et de pouvoir les modéliser avec des techniques automatisées. Par contre, pour obtenir des modèles plus précis, la photogrammétrie et la lasergrammétrie terrestre sont plus adaptées mais demande des temps de traitements plus longs et la taille des modèles produits est plus importante.

Une deuxième utilisation est le LiDAR atmosphérique. Dans ce cas, l'objectif n'est plus de déterminer la position d'éléments composant la surface terrestre mais de détecter, localiser et mesurer la concentration des polluants présents dans l'atmosphère. Ici, ce sont les constituants de l'atmosphère (molécule, aérosol, poussière...) qui vont rétrodiffuser les impulsions lasers. Pour mesurer la concentration d'un polluant, le principe consiste à émettre simultanément un faisceau laser dans le domaine spectral d'absorption du polluant et un deuxième faisceau en-dehors de la zone d'absorption, la concentration du polluant sera fonction de l'écart d'amplitude entre les deux signaux. Le LiDAR est également utilisé pour mesurer la vitesse du vent mesurant l'effet Doppler sur le signal rétrodiffusé. Dans ce TFE, l'étude porte sur la contribution du LiDAR topographique.

Représentations des données LiDAR

Les données LiDAR peuvent être représentées sous la forme d'un nuage de points bruts où chaque point est caractérisé par ses coordonnées x, y, z et complété avec d'autres attributs comme l'intensité et l'écho du signal émis. Pour visualiser ce nuage de points dans les logiciels de SIG (Système d'Informations Géographiques), ce fichier de points au format texte peut être transformé en format vecteur.

L'autre représentation possible des données Lidar est la production d'un modèle numérique d'altitude sous la forme d'une grille régulière (raster) issue de l'interpolation du nuage de points brut. Le pas de la grille (P_g) est choisie en fonction de la densité (P_p) du nuage de points LiDAR. Une relation a été définie (Behan, 2000) pour déterminer la résolution de la grille en se basant sur l'équation suivante : $P_g = 1/\sqrt{D_p}$

Ces deux modes de représentation (vecteur ou raster) ont chacun leurs avantages et leurs inconvénients.

-Le mode vecteur permet d'observer directement les coordonnées (x, y, z) de tous les points avec une grande précision. Mais en raison du volume important de

ces données, la manipulation est très lourde et l'utilisation des algorithmes s'avère très complexe et très longue.

-Le mode raster est facile à utiliser et permet d'utiliser des techniques de traitement d'images et de télédétection. D'un autre côté, l'interpolation du nuage de points LiDAR sous la forme d'une grille génère une perte d'information originale et détériore la précision initiale du modèle issue du nuage de points brut.

Des techniques de classifications peuvent s'appliquer dans ces deux modes et elles vont être présentés dans la suite de cet état de l'art sur les données LiDAR.

Classification basée sur le mode vecteur

La classification peut se réaliser directement sur le nuage de points LiDAR. Le traitement des points LiDAR se déroule en deux étapes : la segmentation pour reconstruire les différentes surfaces interpolées du nuage de points et la classification pour répartir ces surfaces en différentes classes.

Pour réaliser la segmentation, plusieurs méthodes existent :

-*l'algorithme de RANSAC* (RANdom Sample Consensus) : Cette algorithme est utilisée en vision par ordinateur. La méthode consiste, dans un premier temps, à sélectionner des points aléatoirement. On prend un nombre de points minimum pour définir la surface recherchée et on détermine les paramètres du modèle par la méthode des moindres carrés. Dans un second temps, on détermine la pertinence de ce modèle. Pour cela, une marge autour du modèle est appliquée et on compte le nombre de points inclus dans cette marge. Ces points sont définis comme des « inliers ». On réitère ces deux étapes dont le nombre d'itérations n est donné par l'équation (Fischler et Bolles, 1981) $n = \frac{\log(1-p)}{\log(1-w^s)}$ avec p la probabilité d'obtenir

un modèle correct dans des observations avec $w\%$ de mesure fausses. S est le nombre minimal de points nécessaires pour décrire le modèle (3 points pour un plan par exemple). La surface qui possède le plus d'« inliers » est conservé et elle est affiné à la fin en réalisant un ajustement à partir de ces « inliers ».

-*Transformée de Hough* : les points LiDAR sont projetés dans l'espace des paramètres objets. Pour détecter des plans, on constate que les points dans l'espace des paramètres sont des plans et que les points dans l'espace des paramètres correspondent à des plans dans l'espace euclidien. L'objectif de cette méthode est donc de détecter les points dans l'espace des paramètres dont les coordonnées donnent les paramètres des plans ajustant au mieux le nuage de points.

-*méthode par surface mobile* : La méthode par surface mobile combine l'ajustement par moindres carrés et la croissance de surface. L'idée est de créer une grille sur le nuage de points en fonction des coordonnées planes. Sur cette grille, une fenêtre se déplace cellule après cellule. Les points contenus dans cette fenêtre sont utilisés pour calculer les paramètres d'un plan par la méthode des moindres carrés. Les paramètres ainsi que l'erreur moyenne quadratique sont ensuite enregistrés dans la cellule qui est au centre de la fenêtre. Les cellules qui ont les erreurs quadratiques les plus petites sont sélectionnées comme surfaces graines. On compare les paramètres des cellules voisines avec ceux des surfaces graines. Si les cellules voisines ont leurs paramètres proches, elles sont agrégées avec les surfaces graines. On réitère ce processus jusqu'à ce qu'il n'y ai plus de nouvelles surfaces agrégées entre deux itérations.

Pour classifier les surfaces qui ont été interpolées sur le nuage de points LiDAR, les

méthodes principales sont :

-la détection de surface 3D (Lee et Schenk, 2002) : Cette méthode permet de classifier le nuage de points en deux classes qui sont la classe du sol et celle des bâtiments sans tenir compte de la végétation. Dans un premier temps, l'algorithme divise le nuage de points en petits groupes de points, en utilisant le critère de distance entre les points. Un point dans chaque groupe est sélectionné comme représentant une étiquette du groupe. On obtient ainsi deux listes avec les groupes de points et les étiquettes des groupes de points. Dans un second temps, les groupes de points sont assemblés suivant les surfaces que ces groupes représentent. Un point dans chacune des surfaces est sélectionné pour représenter l'étiquette de celle-ci. Suite à cette étape, trois listes sont créées : la première représente les surfaces détectées, la deuxième les contours des surfaces et la dernière les étiquettes des surfaces. À l'issue de ces deux étapes, les bâtiments et le sol sont classifiés simultanément.

-prédiction linéaire (Kraus et Pfeifer, 1998) : Cette méthode part de l'hypothèse que les altitudes des points du terrain sont considérées comme des phénomènes aléatoires. Les points du sursols sont donc considérés comme des erreurs grossières. L'altitude de chacun des points est interpolée à partir de ces points voisins en utilisant une méthode de prédiction linéaire. Elle consiste à interpoler une surface moyenne passant par le nuage de points. Puis on affecte à chaque point un poids correspondant à la distance entre le point considéré et la surface moyenne. Une seconde surface moyenne passant par les points pondérés du nuage est calculée. Cette étape est répétée jusqu'à ce que le calcul se stabilise.

-algorithme MDL (Minimum Description Length) basé sur le critère de longueur de description minimum (Axelsson, 1999) : Les points du sol sont classifiés dans un premier temps en considérant que la surface du sol est liée avec les points bas du nuage de points LiDAR. "La surface peut varier autour d'une certaine valeur en respectant le critère MDL contrainte par une fonction spline et le modèle du contour actif" (CARNEIRO, 2011). Dans un second temps, la végétation et les bâtiments sont classifiés en utilisant le type d'écho. En effet, la particularité du laser à traverser la végétation produit des échos à différentes altitudes, ce qui permet de discriminer ces deux classes. Cette classification est affinée en employant un critère MDL. "Une fonction de coût est définie pour les classes des bâtiments et de la végétation et qui se base sur la dérivée seconde de la différence d'élévation. Le modèle considère que les bâtiments sont considérés comme des surfaces planes liées les unes aux autres et la végétation comme des points avec des dérivées secondes distribuées aléatoirement."(CARNEIRO, 2011)

Passage du mode vecteur au mode raster

Avant d'utiliser les méthodes de classification spécifique au mode raster avec des techniques de traitement d'images et de télédétection, une étape doit être effectuée au préalable. Une interpolation du nuage de points LiDAR est réalisée pour passer du mode vecteur au mode raster. Le nuage de points est transformé en une grille régulière composée de valeurs liées soit à l'élévation, l'intensité ou l'écho. Il existe plusieurs méthodes d'interpolation spatiales pour obtenir cette transformation :

-*méthode de distances inverses* (Shepar, 1968): elle est souvent utilisée pour l'interpolation de points irrégulièrement espacés. Les points Lidars sont pondérés de telle sorte que les points plus éloignés du noeud de la cellule soient moins pris en compte que les points à proximité du noeud.

-*méthode du Krigeage* (Cressie, 1993) : Elle calcule une élévation au centre de la cellule comme étant la moyenne pondérée des élévations mesurées sur les

points LiDAR. Elle se déroule en deux étapes : la détermination du poids et l'estimation de l'élévation en réalisant une moyenne pondérée. "Les poids sont déterminés par la modélisation d'un varigramme qui décrit la variabilité spatiale des valeurs d'élévation autour du point de référence."(CARNEIRO, 2011)

-*création d'un réseau de triangle irrégulier* (Franklin 1973) : un algorithme basée sur la triangulation de Delaunay est appliqué de manière à minimiser les angles de tous les triangles. Les points sont reliés de manière à ce qu'aucun côté du triangle ne soit intersecté par les autres triangles. Une interpolation linéaire est appliquée pour transformer le TIN (Triangular Irregular Network) en image.

-*interpolation par plus proche voisin* : la valeur de la cellule prendra la valeur d'élévation du point LiDAR qui est le plus proche du noeud de la cellule. L'avantage de cette technique est de conserver les valeurs d'altitudes originales et le temps de traitement est rapide.

Lorsque le nuage de points a une densité élevée (au moins un point/m²), les méthodes d'interpolation présentée ci-dessus fournissent des résultats qui n'ont pas de différences significatives (Gonçalves, 2006) .

À l'issue de cette interpolation, on obtient des modèles numériques liés à l'altitude, à l'intensité ou à l'écho suivant l'attribut choisi pour interpoler. Lorsque l'attribut qui a été utilisé pour l'interpolation est une altitude, l'image obtenue est un modèle numérique d'altitude (MNA) sous la forme de raster. Elle représente une matrice à deux dimensions avec des cellules ayant chacune une valeur liée à l'altitude. Lorsque ce modèle représente l'altitude du terrain nu sans infrastructures, il est appelé modèle numérique de terrain (MNT). Lorsque le modèle numérique d'altitude représente les éléments de la surface sans faire de distinction entre les milieux naturels et artificiels, on parle de modèle numérique de surface (MNS). Cette surface comprend la végétation, les arbres, les bâtiments et le terrain nu en l'absence d'autres éléments. À partir du MNS et du MNT, il est possible d'obtenir un modèle numérique de hauteur (MNH) aussi appelé modèle numérique de surface normalisée. Il est calculé en effectuant la différence entre le MNS et le MNT. Il existe également des modèles numériques conçus à partir d'une approche hybride. Cette approche consiste à combiner des modèles numériques d'altitudes interpolés à partir du nuage de points LiDAR avec des données vectorielles 2D des bâtiments. L'intersection du modèle numérique d'altitude sous forme raster et les données vectorielles 2D des bâtiments forme un modèle numérique urbain de surface en 2,5D (MNUS-2,5D). Avec le MNUS-2,5D, il est possible de calculer des critères morphologiques (CARNEIRO, 2011) tel que le volume des bâtiments, la surfaces des façades et des toits en ayant les limites des bâtiments avec les données vectorielles et l'information altimétrique avec le modèle numérique d'altitude sous forme raster.

Le fichier brut de points LiDAR contient non seulement des informations liées à l'altitude mais également à l'intensité et au type d'écho. Ainsi, au lieu de rééchantillonner le nuage de points en utilisant l'information d'élévation, il est également possible de tenir compte de l'intensité et du type d'écho. La classification peut donc également utiliser ces images pour établir des règles de classifications supplémentaire et d'affiner la classification. Une étude a été faite pour distinguer l'asphalte, les arbres, les toits et l'herbe en se basant sur l'intensité (Song et al, 2002). L'intensité des points LiDAR dépend de la surface qui a rétrodiffusé l'onde. L'intensité peut donc donner des indications sur les matériaux qui compose la surface. "Pour un laser de longueur d'onde 0,9 nm, l'asphalte réfléchit l'onde entre 10 à 20%, les toits entre 20 et 30%, les arbres entre 30 et 60% et l'herbe, 50 %" (Song et al, 2002). Pour le type d'écho, la végétation ou les bords de toiture renvoient

plusieurs échos contrairement au sol ou aux pans de toit. Ces informations peuvent donc permettre de distinguer de nouveaux éléments.

Pour réaliser la classification de couverture du sol à partir de données LiDAR, l'utilisation de ces modèles a été utilisée pour distinguer les différentes classes (Antonarakis, 2008).

Les modèles numériques liés à l'élévation, l'intensité ou l'écho sont des images numériques sur lesquelles des techniques de traitement d'images ou de télédétection peuvent être utilisées. C'est pourquoi, la partie suivante passe en revue un état de l'art des différentes techniques de classification qui peuvent s'appliquer sur des images numériques.

Classification basée sur le mode raster

Le premier groupe de méthode consiste à utiliser les techniques de classification d'images numériques normalement réalisées en traitement d'image ou en télédétection. Cette partie détaille les méthodes de traitement d'image d'un point de vue général et dans la partie d'après, l'application de ces méthodes pour le LiDAR sera développée.

Les techniques de traitement d'image peuvent s'appliquer sur des pixels ou sur un ensemble de pixels. Dans ce dernier cas, on parle d'approche orientée objet. Cette technique se décompose en deux étapes. La première étape est la segmentation qui permet de diviser l'image en plusieurs objets. Ces objets sont créés soit suite à l'agglomération d'un ensemble de pixels adjacents qui ont des caractéristiques homogènes (segmentation par régions), soit à l'issue de la détection de discontinuités au niveau des valeurs spectrales à l'intérieur de l'image (segmentation par frontière). La deuxième étape est la classification qui consiste à attribuer à chaque objet une classe. Pour qu'un objet appartienne une classe, il doit remplir des critères que l'utilisateur a lui même définis. Ces critères peuvent être liés aux valeurs spectrales, à la la forme, au voisinage ou à la texture des objets. Dans le cadre de la classification basée sur les pixels, il n' y a pas de segmentation comme étape préalable. La classification consiste à attribuer une classe à chaque pixel à partir des critères liés à la valeur spectrale. L'utilisateur dispose de plusieurs méthodes pour définir ces critères.

•définition des critères par seuillage

L'utilisateur définit des seuils qui donne les critères d'appartenance d'un objet ou d'un pixel à une classe. Les objets ou pixels qui ont des valeurs de critère comprises dans tous les seuils fixés appartiennent à la classe.

•définition des critères par logique floue

L'utilisateur peut également utiliser la logique floue. Le principe consiste à définir des fonctions d'appartenance pour chaque critère et dans chaque classe. Une fonction d'appartenance définit la probabilité d'une valeur d'appartenir à une classe. À partir de ces fonctions, la probabilité d'appartenance de l'objet ou du pixel à la classe est calculée. Si cette valeur est supérieure a une valeur minimum d'appartenance fixée, la décision est prise que cette objet ou pixel appartienne à la classe. L'intérêt de cette méthode est de pouvoir modéliser l'incertitude sur la définition du critère. L'autre intérêt est de donner un poids différent à chacun des critères en appliquant

des fonctions d'appartenances ayant des formes différentes (gaussiennes, linéaires, fonction logistique, triangulaires linéaires, échelon...).

- définition des critères par classification supervisée

La dernière méthode est la classification supervisée. L'utilisateur saisi un ensemble d'échantillons qui correspond aux différentes classes. Ensuite un algorithme va répartir les pixels ou les objets composant l'image en se basant sur les échantillons pour déterminer les valeurs des critères.

Les autres techniques de classifications qui s'appuient sur les méthodes de traitement d'image ou de télédétection est la méthode basée sur un algorithme de croissance de région. La méthode basée sur un algorithme par croissance de région est une technique de traitement d'image basée sur les pixels et ne nécessite pas de segmentation. Le principe consiste à sélectionner le pixel qui a l'élévation la plus petite. Les voisins du pixel sont testés. Pour cela, la dénivelée entre le pixel voisin et le pixel sélectionné est calculée. Si la dénivelée est inférieure à un seuil fixé, le pixel voisin est classé comme "sol". On réitère le test avec les nouveaux pixels classés comme "sol". Cette algorithme s'arrête lorsqu'il n'y a plus de nouveaux pixels classés comme "sol" entre deux itérations. On constate qu'au fur et à mesure de l'avancement de l'algorithme, le nombre de pixel adjacents augmentent.

Cet algorithme de croissance de région est donc intéressant pour classer le sol du sursol.

Le deuxième groupe de méthode pour réaliser la classification sur les images numériques se base sur l'utilisation de morphologie mathématique. Les opérateurs morphologiques les plus utilisés sont l'érosion ou la dilatation.

Les filtres de morphologies mathématiques peuvent être utilisés pour détecter les points du sursol en employant un opérateur d'ouverture sur le MNS.

Le troisième de groupe de méthode est la recherche de plans moyens pour éliminer la végétation de la classe sursol. Cette méthode (Alharty et Bethel, 2002) propose d'utiliser deux échos pour éliminer globalement la végétation de la classe sursol. Afin d'affiner la classification de la végétation, une étape consistant à rechercher les plans moyens permet de discriminer plus précisément la végétation des bâtiments. En effet, les bâtiments sont généralement composés de plans contrairement à la végétation.

La dernière groupe de méthode basée sur les images numériques sont celles utilisant la transformée de Fourier. Le nuage de points est segmenté en utilisant les caractéristiques fréquentielles. Le principe est que les points appartenant à la classe du sol sont caractérisés par des basses fréquences et que les hautes fréquences désignent les objets de la classe sursol. Il faut interpoler le nuage de points pour créer un MNS sous forme raster puis on applique une transformée de Fourier sur cette image originale. Cela permet de calculer deux nouvelles images, une image de basses fréquences et une autre de haute fréquence. L'image filtrée est calculée par l'intermédiaire d'une transformée inverse de Fourier.

c. LANDSAT dans le cadre de l'utilisation des FVM

Pour produire les données d'utilisation du sol pour le modèle FVM, les images LANDSAT sont généralement utilisées. C'était le cas pour l'application du modèle FVM à la ville de Lahore au Pakistan où la classification du mode d'utilisation du sol s'est appuyée sur des images LANDSAT 7 ETM+ (HAFTENDORN 2010). La procédure employée est une classification non supervisée, c'est à dire que l'utilisateur a lui même défini les règles de classification et déterminer les limites entre les classes. La classification a été réalisée avec la technique de la segmentation orientée objet.

Cette technique s'est réalisée en deux étapes. La première est la segmentation qui a pour objet de diviser l'image en plusieurs objets qui ont des caractéristiques homogènes (valeurs spectrales ou formes). La deuxième étape est la classification qui consiste à attribuer à chacun des objets une classe en suivant des règles de classification qui ne sont pas seulement liés à la valeur spectrale des pixels mais également aux valeurs spectrales, à la forme, à la texture ou aux contexte des objets. Les classes créées étaient l'"urbain dense", "urbain peu dense", "végétation", "sable et sol nu" et "eau". Les images LANDSAT 7 ETM+ sont constituées de 9 canaux (bleu, vert, rouge, proche IR, moyen IR, IR thermique "low gain", IR thermique "high gain", IR lointain, panchromatique). Pour cette application, le canal rouge a été utilisée pour discriminer la végétation qui apparaît plus sombre que dans le reste de l'image. Il a également été utilisé pour classifier les zones de sable et de sol nu qui apparaissait plus clair. Le canal infrarouge moyen a permis de faire ressortir l'eau qui apparaissait très foncé sur l'image de ce canal. La texture des objets dans le canal infrarouge thermique a permis de distinguer les zones urbaines denses des zones urbaines peu denses où les zones urbaines denses sont caractérisées par une certaine homogénéité. Ainsi, avec la technique de la segmentation orientée objet et l'utilisation des canaux présentés ci-dessus, il est possible de réaliser une classification sur des images LANDSAT et qui permet de calculer des données d'utilisation du sol utilisables par le modèle FV.

Nous avons passé en revue les différents états de l'art des sujets qui sont traités dans ce mémoire. Maintenant, les méthodes qui s'appuient sur des parties de ces états de l'art et qui ont permis de réaliser les classifications pour obtenir les données d'utilisation du sol utilisables dans le modèle FVM sont présentées dans la partie suivante.

Méthodes utilisées

L'étude consiste à réaliser deux classifications du mode d'utilisation du sol de Genève. Une première classification est basée sur les images LANDSAT 7 ETM+ et une seconde sur les données LiDAR. Les données d'utilisations du sol issues de ces deux classifications sont insérées dans le modèle FVM puis les résultats sont comparés. L'objet de cette partie est de présenter les méthodes qui ont abouti à la réalisation de ces deux classifications ainsi que la méthode utilisée pour calculer les données d'utilisation du sol à partir de ces deux classifications.

a. Méthode de classification des images LANDSAT

Cette partie expose la méthode qui a permis de réaliser la classification du mode d'utilisation du sol à partir d'images LANDSAT 7 ETM+. Cette classification dans la suite du mémoire sera désignée par le terme « classification LANDSAT ».

• zone d'étude

La classification porte sur le milieu urbain genevois et la zone d'étude est centrée sur l'agglomération genevoise. Le relief a des effets importants sur les flux météorologiques, notamment avec les brises des montagnes. La zone étudiée englobe également les montagnes qui entourent la région genevoise (Jura et préalpes).

• Description des données utilisées

Les données utilisées pour réaliser la classification basée sur des images standards sont des données LANDSAT 7 ETM+ (Enhanced Thematic Layer +). Elles sont issues du programme LANDSAT qui est l'un des plus anciens programmes d'observation par satellite de la surface terrestre. Sept satellites ont été lancés depuis 1972. Le dernier en date est le satellite LANDSAT 7 qui a été mis en orbite en 1999. Il est équipé du radiomètre ETM+ qui dispose de 8 bandes de fréquence qui produit des images multispectrales LANDSAT ETM+ composées de 8 canaux (Cf Figure 7). La résolution des canaux est de 30m sauf pour les deux canaux infrarouges thermiques (60m) et le canal panchromatique (15m).

Canal	Bande spectrale	Résolution	Couleur
1	0.45-0.515 μm	30 m	Bleu
2	0.525-0.605 μm	30 m	Vert
3	0.63-0.69 μm	30 m	Rouge
4	0.75-0.90 μm	30 m	Proche IR
5	1.55-1.75 μm	30 m	Moyen IR
6.1	10.4-12.5 μm	60 m	IR Thermique « low gain »
6.2			IR Thermique « high gain »
7	2.09-2.35 μm	30 m	IR Lointain
8	0.52-0.9 μm	15 m	Panchromatique

Figure 7: Description des canaux composant les images LANDSAT 7 ETM+ (HAFTENDORN 2010)

Les images LANDSAT sont téléchargées gratuitement sur le site du Global Landcover Facility (GLCF).

<http://glcfapp.glcg.umd.edu:8080/esdi/index.jsp>

L'image utilisée pour réaliser la classification est présentée ci-dessous.

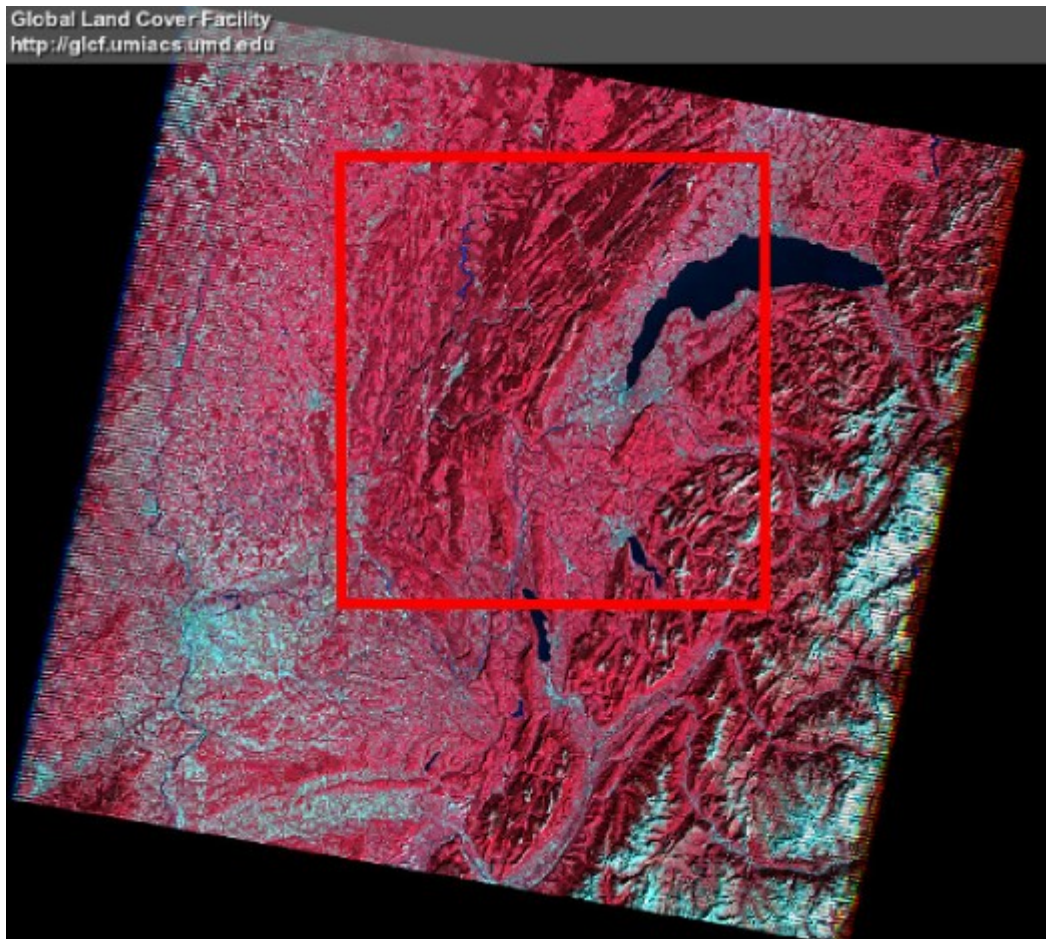


Figure 8: Image utilisée pour la classification avec zone d'étude (en rouge): taille de l'empreinte : 244km*218km ; coordonnées (UTM zone 31N) au centre de l'image : X=721 515m et Y=5 101 185m ; Id 220-860 ; WRS 2 ; Path 196 ; Row 028 ; 21/09/2006; USGS ; Ortho, GLS 2005 ; GeoTIFF ; France, Italie, Suisse

Comme l'image utilisée contient des défauts sur les bords de l'image avec la succession de traits noirs qui n'ont pas de valeurs spectrales (Cf. Figure 9). La zone d'étude est restreinte. Cette diminution de la zone d'étude ne pose pas de problème car Genève et les reliefs qui l'entoure sont correctement représentés.

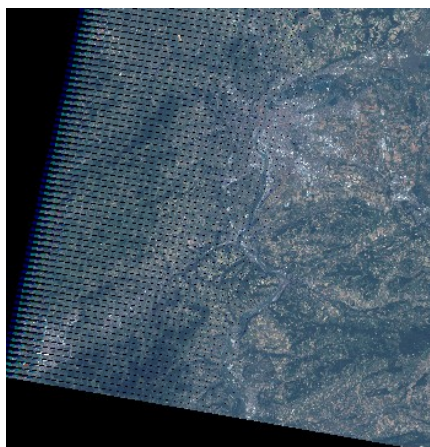


Figure 9: défauts sur le bords de l'image

Au final, l'image sur laquelle la classification LANDSAT va être réalisée représente une région de 105km*105km (7000pixels*7000pixels).

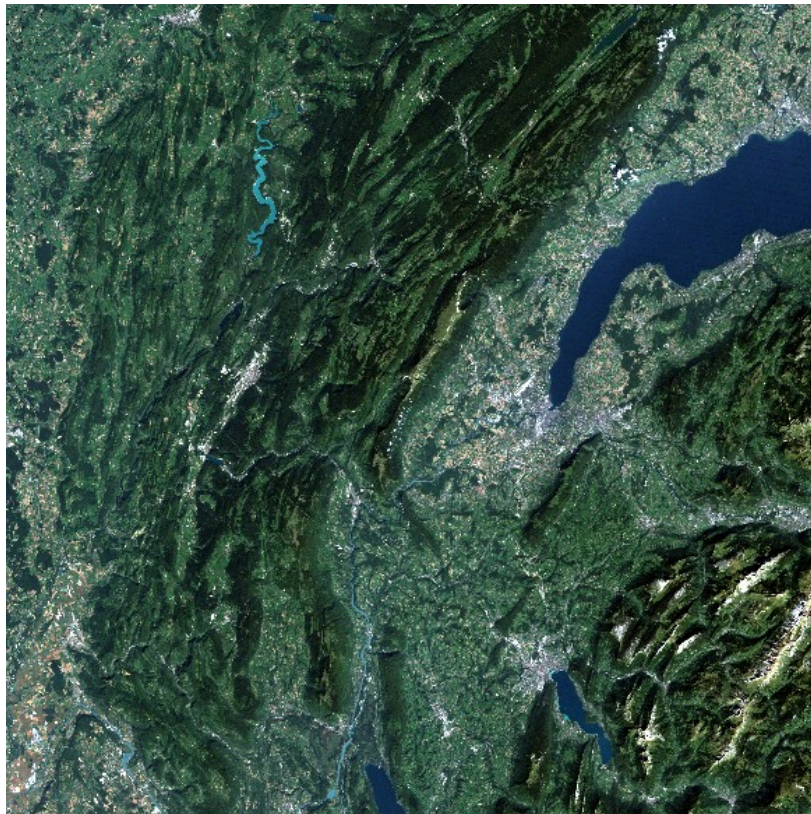


Figure 10: image restreinte sur laquelle la classification va être réalisée (taille de l'empreinte : 105km*105km ; coordonnées (UTM zone 31N) au centre de l'image : X=725 175m Y=5 121 525m

•Description du travail de classification

La classification fut réalisée avec le logiciel eCognition. La procédure de la classification s'est déroulée en deux étapes : la segmentation et la classification. Ces deux étapes ont nécessité d'effectuer plusieurs tests pour déterminer les paramètres optimaux et un résultat satisfaisant. Compte tenu du temps de calcul assez long, la classification s'est appuyée sur une partie de l'image pour déterminer les critères de classification adaptés. Cette partie de l'image contenait un échantillon (Cf Figure 11) représentatif de toutes les classes que nous souhaitons faire ressortir dans ce travail. Une fois les critères déterminés, ils sont appliqués sur la totalité de l'image exploitable.

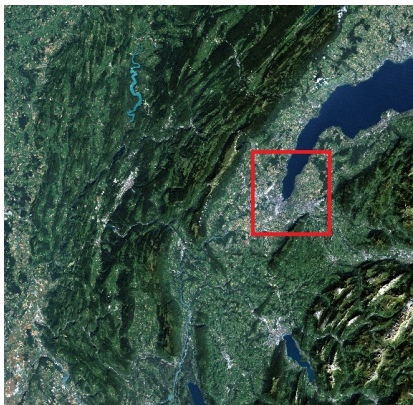


Figure 11: localisation de la zone test dans la scène

La segmentation consiste à diviser l'image en plusieurs régions qui ont des caractéristiques homogènes. L'algorithme utilisée pour réaliser cette étape est la segmentation multirésolution. L'algorithme commence par considérer chaque pixel comme un objet. Ensuite, des paires d'objets adjacents vont être assemblées pour former un plus grand segment. Ce sont ceux qui minimisent l'hétérogénéité qui seront fusionnés. Dès que la valeur minimum d'hétérogénéité dépasse un seuil (l'échelle), la fusion d'un segment avec d'autres objets s'arrête. Pour utiliser la segmentation multirésolution, l'utilisateur doit régler trois paramètres :

-Le paramètre d'échelle : Plus le paramètre d'échelle est grand plus les objets segmentés auront une grande taille.

-L'homogénéité : L'homogénéité est la combinaison de la couleur (valeurs spectrales) et de caractéristiques morphologiques comprenant le *smoothness* (formes irrégulières) et le *compactness* (formes compactes).

-couleur/shape : ce critère caractérise l'influence de la couleur par rapport à l'influence

de la forme. La valeur du shape est établie selon une balance qui s'échelonne de 0,1 à 0,9.

En s'appuyant sur un travail de classification LANDSAT déjà réalisé (Esther Haftendorn, 2010), les valeurs des paramètres retenues sont les suivants :

-poids identiques pour les huit canaux (1)

-échelle : 22

-forme : 0,4

-compacticité : 0,3

Le résultat de la segmentation est présentée sur l'image suivante.



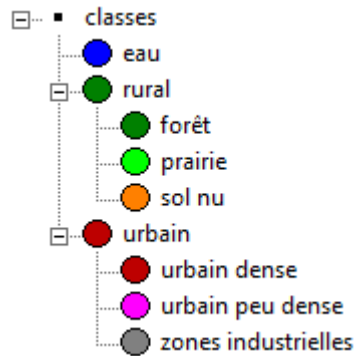
Figure 12: résultat de la segmentation

On constate que les régions qui ont des valeurs spectrales homogènes sont bien délimitées. À partir de cette segmentation, on peut distinguer des objets représentant différentes occupations du sol (eau, forêts, sol nu...). L'objectif de la classification

consistait justement à trouver les règles de classification qui ont permis de distinguer les différentes zones d'utilisation du sol pour calculer les données d'utilisation du sol qui seront insérées dans le modèle FVM.

La méthode de classification utilisée est la classification supervisée en raison de la connaissance du terrain sur la zone en question.. Cette méthode consiste à saisir des échantillons pour chaque classe. Le logiciel calcule automatiquement les règles de classification pour les appliquer sur la totalité de l'image test. Une fois ces règles de classification définies à partir de l'image test, elles sont exportées pour être utilisées sur la totalité de la scène.

En se basant sur les besoins du modèle météorologique ainsi qu'en analysant les possibilités d'objets en différentes classes, les classes suivantes ont ainsi été définies :



prairie



sol nu



eau



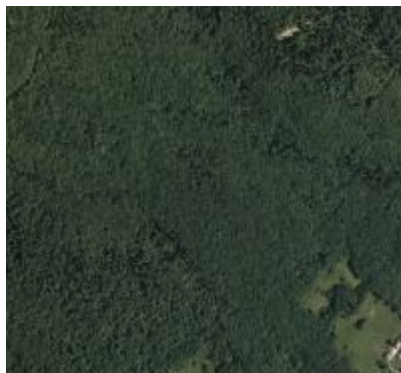
urbain peu dense



urbain dense



zone industrielle



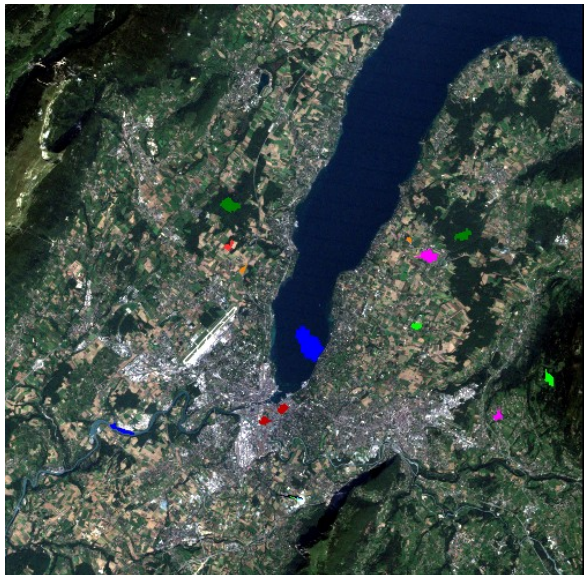
forêt

L'image est constituée d'ombre dans les zones adjacentes aux montagnes ayant un relief escarpé. De plus, l'image est constituée de nuages et d'artefacts. Donc, des classes supplémentaires sont créées pour détecter ces effets indésirables puis une classification manuelle est réalisée pour ces objets classifiés en tant que tels.

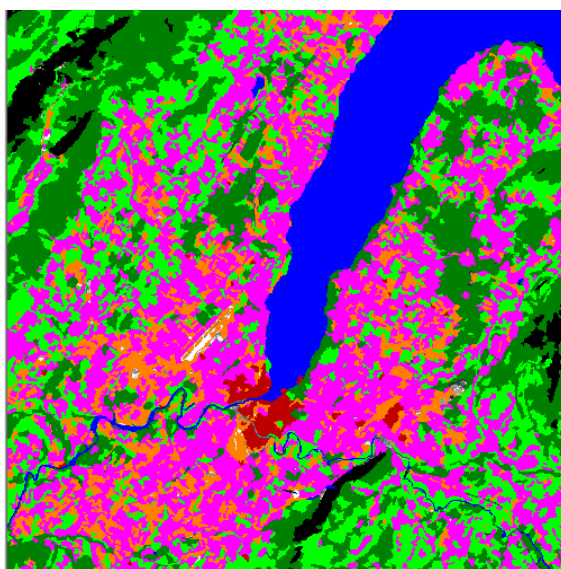
Pour distinguer la végétation (forêt, prairie, terrain cultivé), l'indice NDVI est calculé pour chaque pixel. Une image est créée en utilisant les canaux Proche IR, et Rouge puis en appliquant la formule $NDVI = \frac{\text{Proche IR} - \text{rouge}}{\text{Proche IR} + \text{rouge}}$.

Pour la classification supervisée, les critères utilisés sont les valeurs spectrales des 8 canaux et de l'image NDVI. En plus, avec l'approche orientée objet, il est possible d'avoir des informations sur la texture et la densité des objets qui sont calculés automatiquement avec le logiciel eCognition et ne nécessite donc pas de produire une image liée à ces deux critères. Un objet qui a une densité faible est un objet qui a une forme allongée. Ce critère est utile pour distinguer l'eau qui compose les cours d'eau et qui ont une forme étirée. Le critère de texture permet de distinguer les zones urbaines denses des zones urbaines peu denses, car l'urbain dense présente une certaine homogénéité. En plus des valeurs spectrales des 8 canaux et de l'image NDVI, les critères de densité et de texture sont donc également utilisés pour la classification supervisée qui s'est déroulée de la manière suivante.

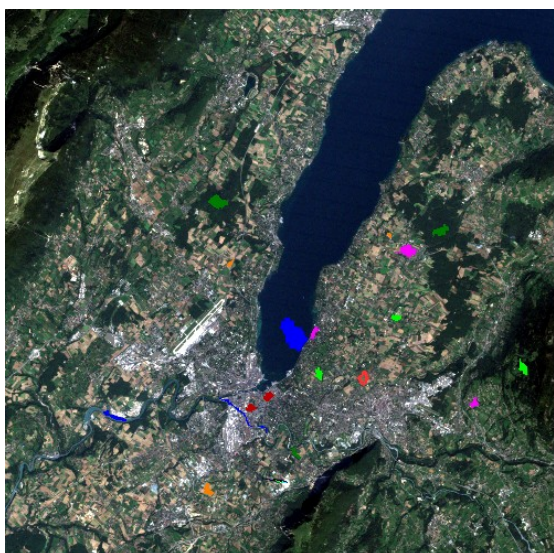
On commence par sélectionner deux échantillons pour chaque classe. Le logiciel calcule automatiquement la valeur des critères adaptés pour chaque classe. On lance la classification. On sélectionne à nouveau pour chaque classe deux échantillons mal classés. Ces opérations sont réitérées jusqu'à ce que le résultat soit satisfaisant. Trois itérations (sélection et échantillons) sont nécessaires.



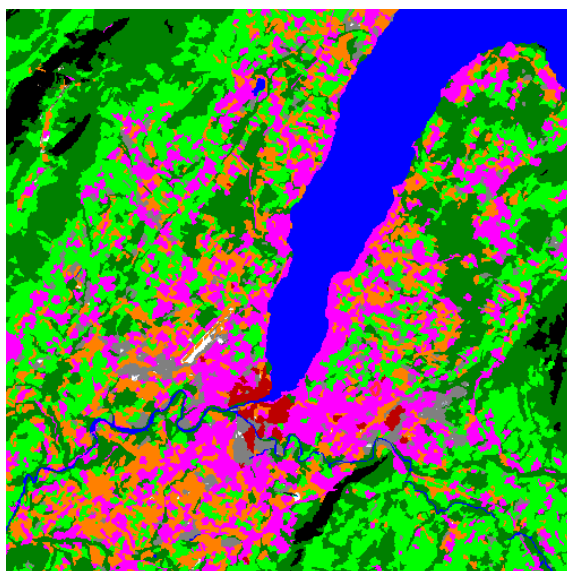
sélection des échantillons



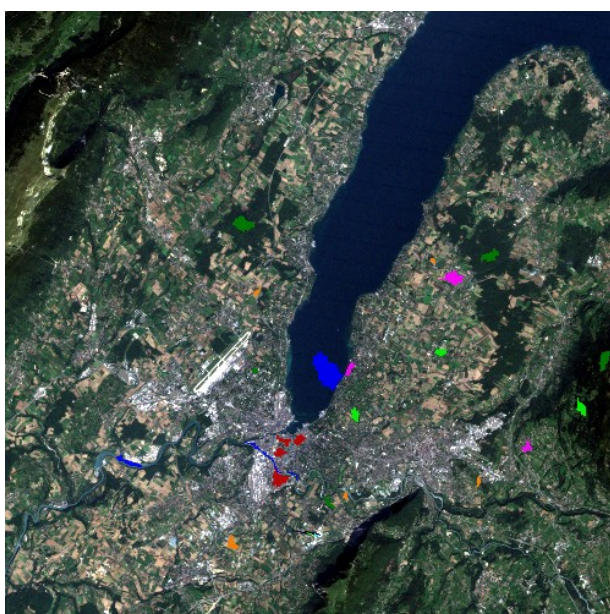
résultat issu de la première classification



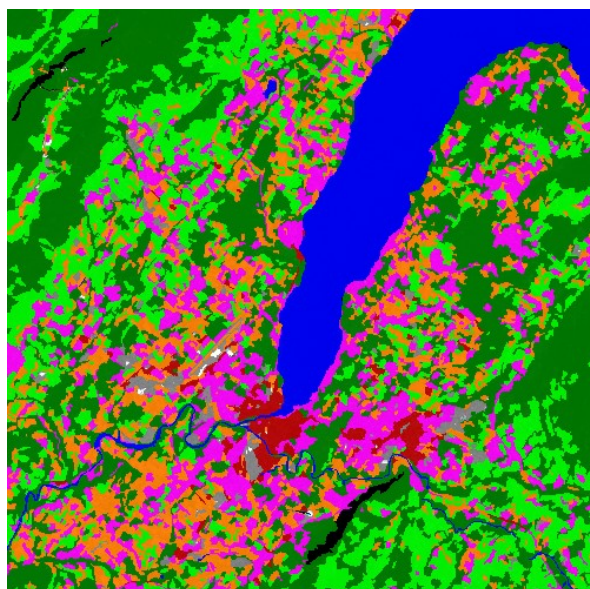
deuxième sélection d'échantillons



résultat de la deuxième classification



3ème sélection d'échantillons



résultat de la 3ème classification

Ces échantillons sont exportés pour utiliser ces propriétés pour la classification sur la totalité de la scène. Après avoir classifié manuellement les effets indésirables présents sur l'image (ombres, artefact et nuage), on obtient le résultat suivant :

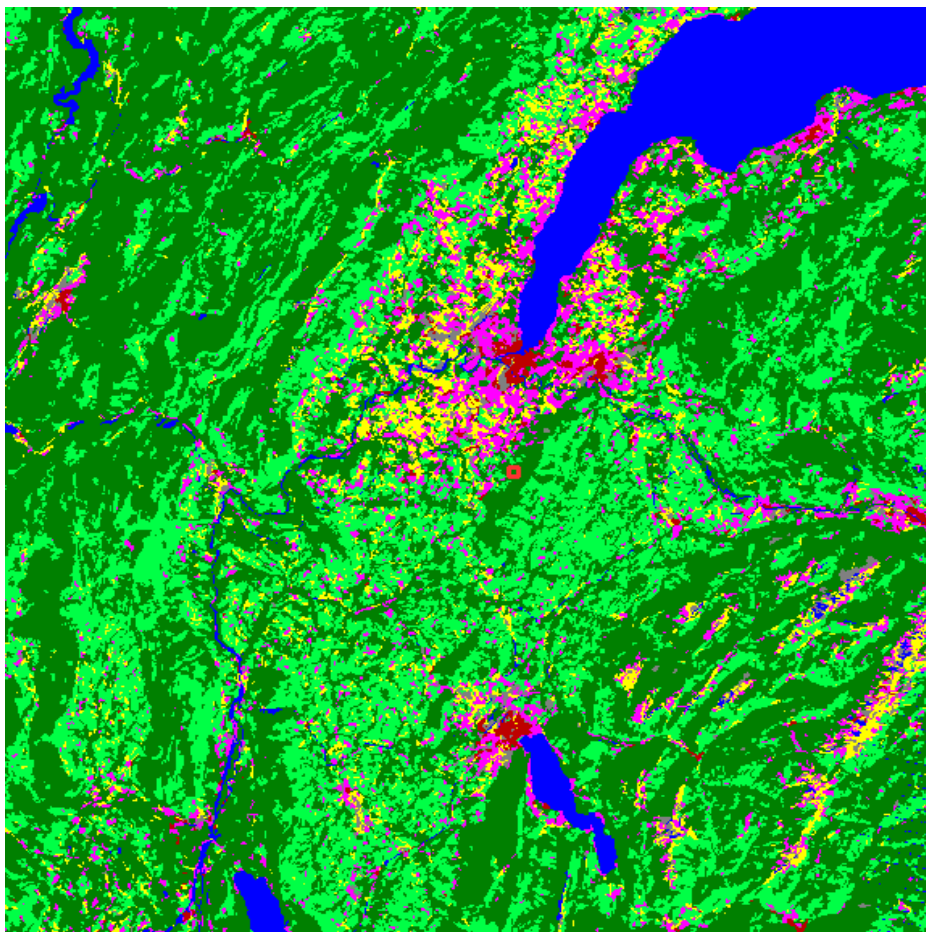


Figure 13: résultat de la classification supervisée sur la scène entière

Pour obtenir les mêmes classes que celles présentes dans la classification Lidar, les classes d'occupation du sol « forêts, prairie et sol nu » sont agrégées pour former la classe d'utilisation du sol « rural ». Cela ne pose pas de problème car le modèle FVM applique la paramétrisation rurale sans faire de distinction entre ces différentes classes. De plus, la classe « zone industrielle » est fusionnée avec la classe « urbain peu dense ». En effet, les zones industrielles sont constituées de bâtiments avec une faible hauteur et très espacés entre eux.

Le résultat final de cette classification est présentée dans la partie « Résultats ».

b. Méthode de classification des données LiDAR

Cette partie présente la méthode qui a permis de réaliser la classification du mode d'utilisation du sol à partir des données LiDAR combinées avec la carte 2D du bâti.

La méthode choisie consiste à réaliser une classification basée sur des images numériques. Ces images numériques sont issues de l'interpolation du nuage de points LiDAR. Chaque point LiDAR contient des informations sur ses

coordonnées, l'intensité et le type d'écho. Ainsi, après interpolation, on peut obtenir des images basées sur ces informations et y appliquer les techniques de traitement d'image et de télédétection pour produire la classification.

Zone d'étude

Les données LiDAR couvre une zone plus restreinte que les images acquises par le satellite LANDSAT 7. De plus, la précision des données LiDAR demande un temps de traitement plus long que sur les images satellites. En effet, il faut réaliser des algorithmes de traitements sur le nuage de points LiDAR ou réaliser une interpolation de ce nuage pour obtenir des informations sous la forme d'une image. Il a donc fallu sélectionner une zone d'étude qui couvre un territoire moins étendue que celui de l'image LANDSAT mais qui permettent de mettre en évidence de potentielles différences une fois que les données d'utilisation du sol issues du LiDAR et issues des images LANDSAT sont insérées dans le modèle FVM. C'est pourquoi, la zone d'étude doit être suffisamment grande de manière à ce qu'elle occupe une zone rurale, une zone urbaine et la transition entre les deux. La zone d'étude est présentée ci-dessous et représente une zone de dimension 8km*5km.



Figure 14: localisation de la zone d'étude de la classification issue du LiDAR

Description des données utilisées

La classification s'est appuyée sur la combinaison de plusieurs sources de données : un modèle numérique de surface sous forme raster (MNS), un modèle numérique de terrain sous forme raster (MNT), un nuage brut de points LiDAR et une carte 2D du bâti.

Le nuage de points LiDAR (Cf Figure 15) existe sous la forme d'un fichier de type XYZ (données attributaires) qui a été transformé en « shapefile » pour être visualisé dans des logiciels de SIG. Il couvre tout le canton de Genève et est constitué d'environ 2500 millions de points. Le nuage de points est divisé en 481 tuiles comprenant chacune une zone de 1 km². Ces données LiDAR multi échos ont été acquises à l'issue d'une campagne aérienne utilisant le système LiDAR aéroporté et réalisée pendant l'année 2009 à l'initiative du système d'information du territoire

genevois (SITG). La densité moyenne des points LiDAR était de 6 points par m². Chaque point LiDAR a des attributs liés à ses coordonnées, le type d'écho, l'intensité. En ce qui concerne le type d'écho, il en existe 4 catégories : écho seul, premier écho parmi plusieurs, écho intermédiaire parmi plusieurs et dernier écho parmi plusieurs échos. Cette information permet de distinguer la végétation qui a la particularité de renvoyer plusieurs échos. L'intensité constitue la mesure de l'intensité du signal reçu par le système LiDAR.

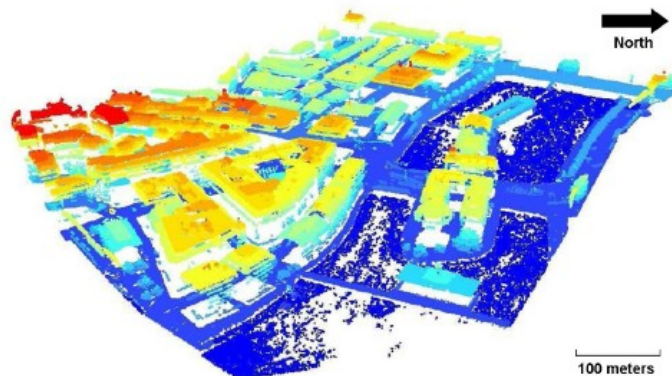


Figure 15: nuage de points LiDAR d'un quartier de Genève (CARNEIRO)

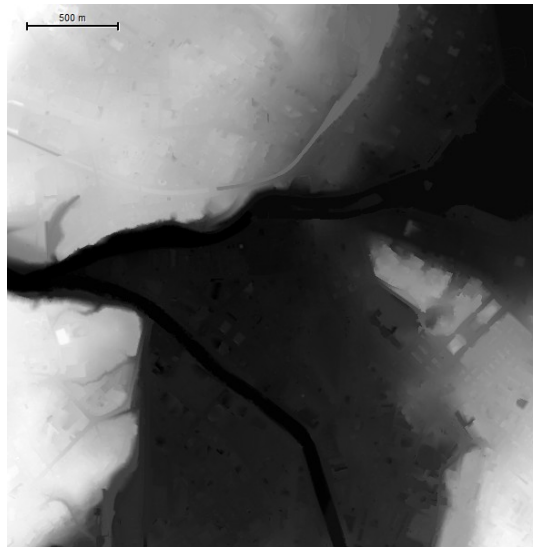
Le modèle numérique de surface (Cf Figure 16) a été produit par le SITG en produisant un réseau de triangle irrégulier (TIN) sur le nuage de points LiDAR. Puis ce TIN a été interpolée linéairement pour obtenir un MNS sous la forme d'une image raster. Le modèle numérique de surface a une résolution de 1 m. Cette résolution est suffisante pour une telle application. En effet, Le MNS issu du LiDAR a une résolution 30 fois meilleure que les images LANDSAT généralement utilisées pour le modèle FVM et il est possible de distinguer des éléments (bâtiments, arbres, sol) que l'on repère difficilement sur des données LANDSAT. Cette différence est intéressante car elle permettra de savoir si les détails classifiés sur le MNS issue de données Lidar contribuent à influencer les résultats qui sortent du modèle.



*Figure 16: MNS du centre ville de Genève (résolution : 1m ; dimension : 3km*3km)*

Le modèle numérique de terrain a également une résolution de 1m pour les

mêmes raisons que le MNS. Le MNT a été conçu par le SITG en effectuant une classification du sol sur le nuage de points LiDAR à partir de l'algorithme MDL. Un réseau de triangle irrégulier (TIN) a été construit à partir de points appartenant à la classe « sol ». Ce TIN a été interpolé linéairement pour obtenir un modèle numérique de terrain sous forme d'une image raster (cf. Figure 17).



*Figure 17: MNT du centre ville de Genève (résolution : 1m ; dimension : 3km*3km)*

La carte 2D du bâti correspond à une projection des toits issues des données vecteurs du modèle 3D du canton de Genève (cf. Figure 18). Ces données ont été acquises par digitalisation d'un modèle stéréoscopique d'images aériennes. Les données vectorielles des toits ont été choisies plutôt que les données de l'emprise des bâtiments car les mesures effectuées par un système aéroporté se font uniquement sur les toits de bâtiments qui cachent les limites du bâtiment au sol.



Figure 18: carte 2D du bâti du centre ville de Genève

Choix des classes

Pour réaliser une véritable comparaison, la classification LiDAR doit avoir les mêmes classes d'occupation que celles présentes dans la classification LANDSAT 7 : urbain dense, urbain peu dense, et rural. Comme les images issues des données LiDAR et les images LANDSAT ont des résolutions différentes, on ne détecte pas les

mêmes informations. Par exemple, sur les images LiDAR, on distingue précisément les bâtiments, le sol et la végétation tandis que sur les images LANDSAT, on perçoit les zones avec du bâti denses ou peu denses, les prairies, les forêts ou les cours d'eau. La difficulté consiste donc à trouver une méthode qui permette de définir les mêmes classes que celles présentes sur les images LANDSAT à partir des éléments que l'on distingue sur les données issues du LiDAR.

Pour aboutir à la classification de ces éléments (urbain dense, urbain peu dense et rural), une classification du mode d'occupation du sol est réalisée en utilisant les classes : eau, bâtiment, végétation, sol perméable, sol imperméable. Puis une méthode a permis d'utiliser ces classes d'occupation du sol pour créer une seconde classification constituées cette fois-ci de classes d'utilisation du sol et qui sont identiques à celles utilisées sur les images LANDSAT.

Classification du mode d'occupation du sol

Les étapes qui ont abouties à la réalisation de la classification du mode d'occupation du sol en utilisant le logiciel eCognition sont détaillées dans cette sous-partie en prenant un extrait de la zone à classifier pour mieux visualiser les étapes. Cet extrait est une tuile représentant une zone de 3km de côté et situé au sud-est de la zone d'étude de la classification issue du LiDAR.

Création du MNH et du modèle d'intensité :

La première étape pour réaliser la classification d'occupation du sol consiste à créer deux images numériques. Le modèle numérique de hauteur (MNH) est produit en calculant la différence entre le MNS et le MNT. L'image numérique d'intensité est obtenue en interpolant le nuage de points avec une résolution de 1m. Un filtre médian est appliquée sur cette image pour réduire l'effet « poivre-sel ».



Figure 19: $MNH = MNS - MNT$



Figure 20: modèle numérique des *intensités*

Classification de l'eau, du sol, des bâtiments et de la végétation :

La classification de ces éléments est basée sur la valeur spectrale des pixels issus du MNS et MNI.

Pour classifier l'eau, les pixels qui ont une hauteur (MNT) valant 0 et qui ont une intensité nulle sont définis comme appartenant à la classe "eau". (cf. Figure 20).

Pour le sol, les pixels non classifiés et qui ont une hauteur inférieure à 1,50m appartiennent à la classe "sol". (cf. Figure 21).

Les pixels qui ont une hauteur supérieure à 1 m sont classés comme « bâtiment ». Or la végétation a la même caractéristique. Ainsi, pour faire la distinction entre la végétation et les bâtiments, les données vectorielles des toits sont utilisées. La superposition des pixels avec la couche vectorielle des toits sont considérés comme appartenant à la classe "bâtiment"(cf. Figure 22). Un modèle numérique de surface urbain normalisé est crée et permet d'obtenir des critères morphologiques sur les bâtiments comme la hauteur moyenne, le volume, la surface des façades.

En classifiant les bâtiments de cette manière, on peut ainsi distinguer la végétation en classifiant tous les pixels qui ont une hauteur supérieure à 1m et qui n'appartiennent pas à la classe "bâtiment"(cf. Figure 23).



Figure 21: classification de l'eau (MNH=0 et Intensité=0)



Figure 22: classification du sol (non classifié et MNH<1,5m)



Figure 23: classification des **bâtiments** (pixels qui superposent les vecteurs toits)



Figure 24: classification de la **végétation** (non classifié et MNH>=1,5m)

Classification des ponts en sol :

La végétation est classée de telle sorte à ce que tous les pixels qui ont une hauteur supérieure à 1m et qui n'appartiennent pas aux bâtiments soient intégrés dans la classe végétation. Or les ponts ont la même particularité. Ils sont également classifiés comme végétation, ce qui n'est pas correct. Par conséquent, une chaîne d'algorithmes a été créée sur eCognition pour faire la distinction. Il part du principe que les ponts sont à la même hauteur que le sol tandis que la végétation est au-dessus du sol. Donc les pixels "végétation" ayant pour voisin des pixels "sol" et qui ont un écart d'altitude (MNS) entre les pixels voisins "sol" et le pixel "végétation" inférieur à 1 m sont classifiés comme "sol". On réitère cet algorithme en tenant compte cette fois-ci des pixels qui sont devenus 'sol'. On effectue cette boucle jusqu'à ce que le calcul se stabilise, c'est-à-dire jusqu'à ce que le nombre de pixels classés "sol" soit identique entre deux itérations.

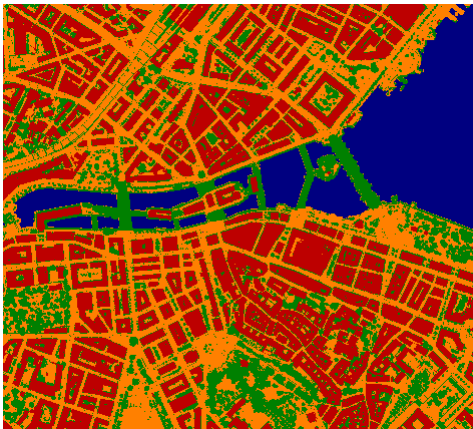


Figure 25: ponts classifiés en végétation

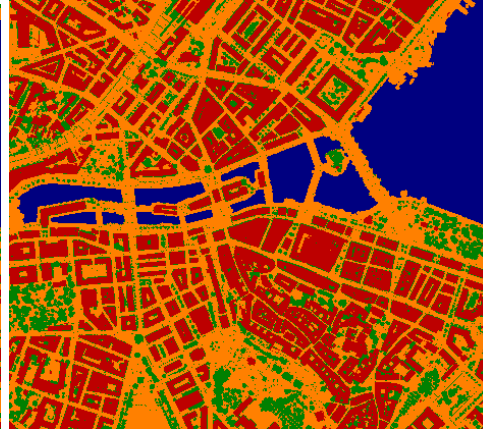


Figure 26: classification des ponts en sol

classification du sol perméable et imperméable

Pour faire la sous-classification du sol en "sol perméable" et "sol imperméable", le modèle numérique d'intensité est utilisé. Les sols imperméables correspondent aux éléments artificiels situés au niveau du sol. Cela comprend les routes, l'asphalte, les surfaces bétonnées. Les sols perméables désignent les éléments naturels qui composent les champs, les prairies ou les surfaces contenant de la végétation basse. Pour les distinguer, les pixels appartenant à la classe "sol" et qui ont une intensité inférieure à 30 sont classifiés comme "sol perméable" et ceux qui sont supérieurs ou égales à 30 comme "sol imperméable".



Figure 27: classification des sols imperméables en gris (sol avec intensité >30) et des sols perméables en vert clair (intensité ≤30)

À l'issue de cette étape, on obtient une classification d'occupation du sol du milieu urbain de Genève avec la représentation des bâtiments, de la végétation, des sols perméables et imperméables ainsi que de l'eau. Le résultat de cette classification du mode d'occupation du sol est présenté dans la partie "Résultats".

Classification du mode d'utilisation du sol

Cette partie présente les étapes qui permettent d'obtenir les mêmes classes d'utilisation du sol que celles présentes dans la classification LANDSAT 7 : urbain dense, urbain peu dense, et rural. Elle montre comment la classification du mode d'occupation du sol est utilisée pour former une classification d'occupation du sol qui soit compatible avec le modèle FVM et qui permettent d'obtenir des données d'utilisation du sol comparables avec celles obtenues à partir de la classification issue des images LANDSAT.

classification des zones urbaines denses et peu denses

La notion de densité dépend de la hauteur des bâtiments et de la proximité des bâtiments entre eux. Plus une zone sera dense, plus les bâtiments auront une hauteur moyenne élevée et la distance entre les bâtiments sera faible. La méthode pour classer les zones denses et les zones peu denses tient compte de ces deux critères.

Pour cela, les bâtiments de la classification d'occupation du sol décrite dans la partie précédente sont classifiés en deux sous-classes (Cf Figure 28) : les bâtiments inférieurs à 15 m et les bâtiments supérieurs à 15m. Ce seuil de 15 m a été choisi en se basant sur la distribution de la hauteur des bâtiments à Genève. Cette distribution a une forme bi-modale (Cf Figure 27) avec deux pics de 8 m et 21 m. Le premier mode situé entre 0 et 15m correspond aux zones résidentielles avec les maisons individuelles et les zones industrielles avec en particulier les hangars. Elles sont situées en périphérie. Le deuxième mode s'étend au delà de 15m et correspond à des immeubles essentiellement situés au centre ville où le bâti est plus dense. C'est la raison pour laquelle, la hauteur de bâtiment a été choisie comme un des critères

pour caractériser la densité.

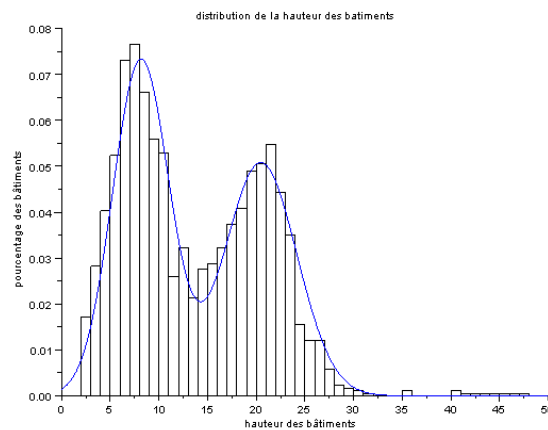


Figure 28: histogramme de la hauteur des bâtiments et ajustement de cette distribution(en bleu)

Les bâtiments ayant une hauteur moyenne supérieure à 15m sont classifiés comme appartenant à la classe "zone urbaine dense". Puis un algorithme de croissance de région est utilisé pour obtenir un étalement de la "zone urbaine dense" sur les classes "sol imperméable" et "végétation" (Cf Figure 29). La croissance de la surface élémentaire de la classe "urbaine dense" est réalisée sur les classes "sol imperméable" et "végétation" car en zone dense, les bâtiments sont séparés par des rues composés d'asphaltes et d'arbres. Pour cela, le principe consiste à classier comme "urbain dense" les pixels appartenant à la classe "sol imperméable" ou "végétation" et qui touchent un pixel "urbain dense". On réitère cette algorithme en se basant sur les pixels nouvellement classifiés. Pour tenir compte de la proximité des bâtiments dans le cas des zones urbaines denses, le nombre d'itération a été limité à 20. En effet, dans les zones denses, on peut considérer que les bâtiments ne sont pas espacés de plus de 20m, ce qui correspond à la largeur d'un boulevard.

Pour classier la classe "zone urbaine peu dense", le principe utilisé est le même que pour la classe "zone urbaine dense". Les critères de hauteur des bâtiments et de distance entre les bâtiments sont utilisés. En ce qui concerne la hauteur, les batiments qui ont une hauteur inférieure à 15m sont considérés comme appartenant à la classe "zone urbaine peu dense". Cette valeur est choisie en se basant sur la distribution des hauteurs des bâtiments à Genève comme expliqué ci-dessus (Cf Figure 27). Pour la distance entre les bâtiments, elle sera plus grande que pour le cas de la classe "zone urbaine dense". Pour déterminer sa valeur, les particularités de la zone résidentielle sont utilisées car la zone résidentielle est caractéristique des espaces peu denses. La zone résidentielle est constituée de maisons individuelles séparées par des jardins et des rues. La distance choisie correspond à la distance de ces deux éléments juxtaposés, soit 35m. En utilisant ces deux critères, un algorithme de croissance de région est utilisé. On obtient un étalement de la "zone urbaine peu dense" sur les classes "sol imperméable", "sol perméable" et "végétation". L'étalement s'est réalisé sur ces trois classes car une zone résidentielle est constituée de rues, de jardins et d'arbres et arbustes.

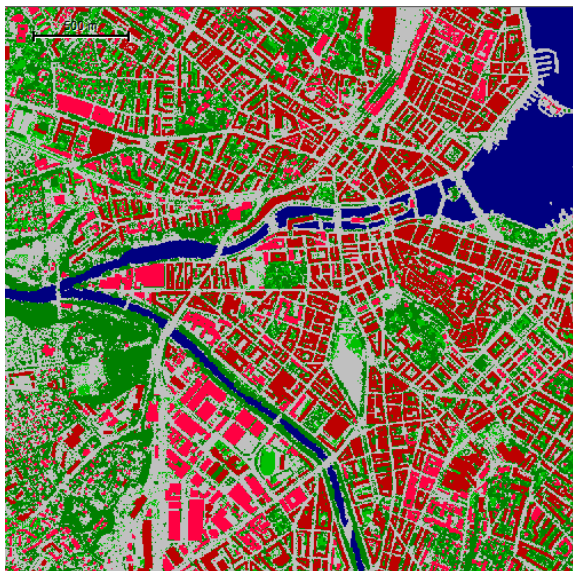


Figure 29: classification des bâtiments inférieurs à 15m (rose) et bâtiments supérieurs à 15m (rouge)

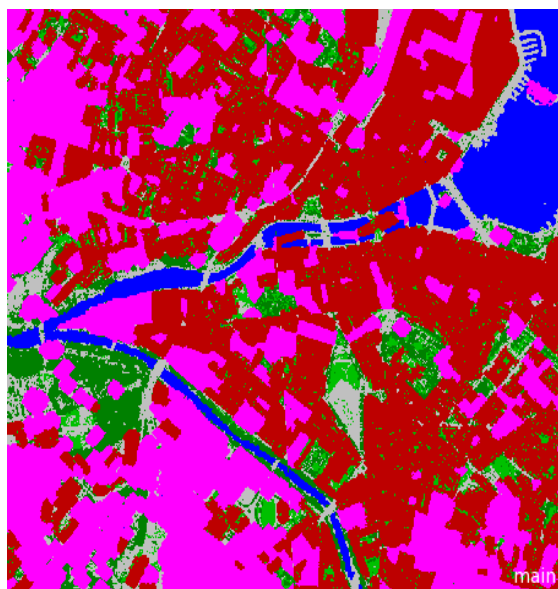


Figure 30: classification de l'urbain dense (rouge) et l'urbain peu dense (rose)

classification des zones rurales

Les zones rurales représentent les forêts avec les zones agricoles composées de champs, sols nus et de bâtiments agricoles. Pour tenir compte des bâtiments agricoles, il a été considéré que les zones urbaines peu denses qui ont une superficie faible correspondant à la dimension de ce type de bâtiments et qui sont entourés d'objets "rural" appartiennent également à la classe "rural". La valeur de la superficie choisie est déterminée en mesurant sur des images aériennes issues du géoportail la superficie d'un bâtiment agricole présent dans le canton de Genève. Sa valeur mesurée est de 2000m². Par conséquent, les objets "urbain peu dense" qui ont une superficie inférieure à 2000 pixels et entourés d'objets "végétation" et "sol perméable" sont classés en "rural" (Cf Figure 31 et 32). Ensuite, comme la végétation et les sols perméables (champs, sol nu, terrain labourée..) sont présents plus fréquemment dans des zones rurales, les classes "végétation" et "zone perméable" sont agrégées pour compléter la classe d'utilisation du sol "rural" (Cf Figure 33).



Figure 31: bâtiments agricoles en urbain peu dense

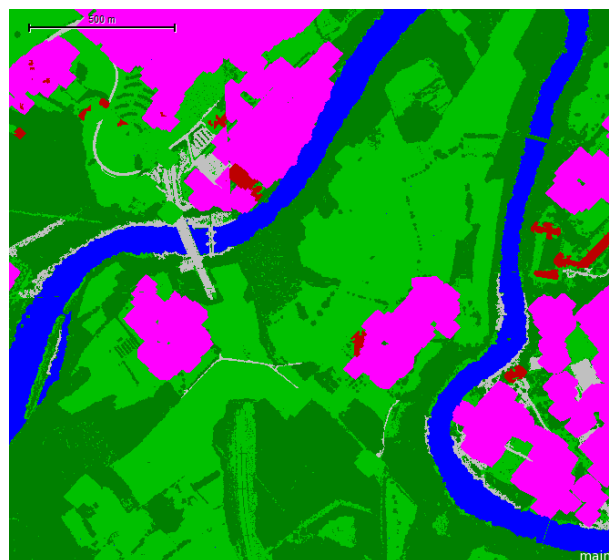


Figure 32: classification des bâtiments agricoles en sol perméable.

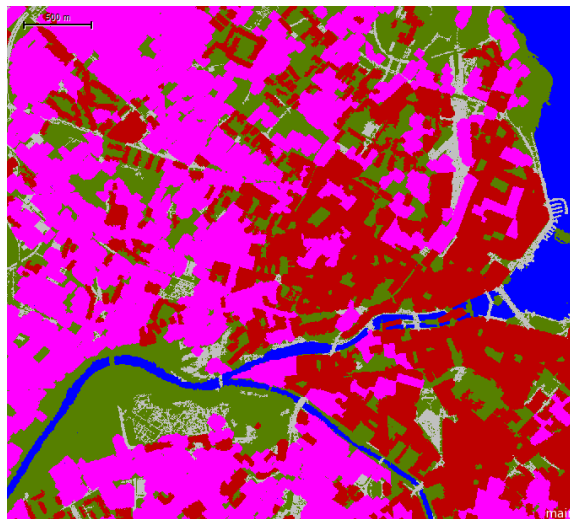


Figure 33: agrégation des classes "végétation" (vert foncé) et "sol perméable" (vert clair) en classe "rural" (vert foncé)

On remarque qu'il ya des objets classifiés comme "rural" en zones urbaines (Cf Figure 33). Cela est du au fait que la végétation et la surface perméable ont été agrégés en classe "rural". Or il y a également ces deux éléments en milieu urbain avec des parcs et des arbres. Pour en tenir compte, les objets "rural" qui ont une superficie inférieure à la taille d'un parc et qui sont en contact avec des objets urbains denses ou peu denses ont été classifiés comme objet "urbain peu dense". La taille retenu est de 85 000m² qui correspond à un parc de Genève (Eaux-Vives).



Figure 34: classification des surfaces imperméables, parcs et jardins en classe "urbain peu dense"

Les zones rurales sont composées également de routes. Pour en tenir compte, les objets linéaires (utilisation de critère de densité < 1) et qui ne touche pas un objet urbain est considéré comme rural.

Pour résumer, cette classification d'utilisation du sol s'est basée sur la classification d'occupation du sol réalisée à partir des données LiDAR et d'une carte 2D du bâti où les bâtiments, la végétation, le sol perméable, le sol imperméable et l'eau étaient représentés. Pour obtenir la classe « urbain dense » et la « classe urbain peu dense », la distance entre les bâtiments et la hauteur des bâtiments ont été prises en compte. Il a également été considéré que les zones urbaines denses sont composées de végétation avec la présence d'arbres dans les rues. Les zones urbaines peu denses sont composées de végétation et de surfaces imperméables avec la présence de parcs qui sont constitués de ces deux éléments. Pour obtenir la classe « rural », il a été considéré que cette classe est caractérisée par des surfaces imperméables avec les prairies, le sol nu, les champs et est également caractérisé par de la végétation avec les forêts. Les classes « végétation » et « sol imperméable » ont ensuite été agrégées pour former la classe d'utilisation du sol « rural ». Pour finir, il a été tenu compte que dans les zones rurales, des routes et des bâtiments agricoles sont présents.

Ainsi, la classification LiDAR a permis d'obtenir au final les mêmes classes d'utilisation du sol que celles présentes dans la classification LANDSAT ETM+, ce qui rend possible une comparaison entre ces deux classifications. Cette partie a expliqué la méthode appliquée à la classification issue du LiDAR sur une zone restreinte (dimension 3km*3km) pour visualiser l'évolution de la réalisation de la classification. Pour obtenir les résultats de cette classification finale sur la totalité de la zone d'étude (8km*5km), elle est présentée dans la partie « Résultats ».

c. Méthode pour calculer les données d'utilisation du sol LiDAR et LANDSAT

Les données d'utilisation du sol issues des classifications doivent être mises en forme de manière à pouvoir les insérer dans le modèle. Pour cela, une grille est superposée sur chacune des classifications et les pourcentages des classes dans chaque cellule de 1km*1km constituent les données d'utilisation du sol. La grille est superposée aux classifications en utilisant eCognition. En effet, ce logiciel permet d'effectuer une segmentation multi-niveaux. La segmentation multi-niveaux permet de réaliser des classifications sur plusieurs niveaux et d'obtenir des informations sur les relations des objets entre les différents niveaux. Par exemple, on réalise une classification sur un niveau constitué d'objets de petites tailles appelés sous objets. On réalise une classification sur un deuxième niveau constitué d'objets de plus grande taille appelés super objets. Avec la segmentation multi-niveaux, on obtiendra des informations sur le nombre de sous objets présents dans un super objet, la surface moyenne des sous objets à l'intérieur d'un super objet ou également le pourcentage de sous objets appartenant à une classe à l'intérieur d'un super objet. C'est cette dernière information qui est utilisée pour calculer les données d'utilisation du sol pour le modèle FVM.

Dans notre cas, la classification LiDAR est réalisée sur un premier niveau. Puis une segmentation en damier (nommé "chessboard segmentation" sur eCognition) est effectuée sur un deuxième niveau où l'on obtenait des objets de formes carrées avec une taille de 1000pixels*1000pixels. Au final, on obtient le pourcentage surfaciques des classes d'utilisation du sol à l'intérieur de chaque objet carré qui constitue les cellules de notre grille.

On a effectué la même procédure pour la classification LANDSAT sauf qu'il a fallu réaliser une opération supplémentaire en raison de la résolution des images LANDSAT qui ont une résolution de 30m. En effet, pour obtenir une grille avec des cellules représentant des cotés de 1000m*1000m, il aurait fallu réaliser une segmentation en damier avec une échelle de $33,33333(=1000/30)$ pixels. Or l'unité de base d'une segmentation est le pixel et il n'est pas possible de travailler sur une portion de pixel. Pour éviter ce problème, la classification LANDSAT a été rééchantillonnée pour que la classification qui a été exportée en format raster ait une résolution de 5m. Ainsi, il est possible de créer une grille avec des cellules ayant un pas de 1000m en réalisant une segmentation en damier avec une échelle de 200 pixels.

Les résultats obtenus sont exportés au format texte où l'on avait les coordonnées des cellules (lignes et colonnes) et les pourcentages surfaciques de chaque classes dans chacune des cellules.

L'étendue de la grille superposée à la classification LiDAR (« grille LiDAR ») constitue une fraction de la grille superposée à la classification LANDSAT (« grille LANDSAT »). En effet, la « grille LANDSAT » est constituée de 105*105 cellules alors que la grille « LiDAR » est composée de 8*5 cellules. On a donc localisé sur la « grille LANDSAT » les cellules correspondant à celles de la « grille LiDAR ». Puis on a remplacé dans le fichier texte les pourcentages surfaciques de ces 40 cellules par les valeurs issues de la « grille LiDAR ». Ce fichier ainsi modifié est considéré comme les « données d'utilisation issues du LiDAR ». Le fichier sans ces modifications constituait les « données d'utilisation du sol LANDSAT ».

Les données d'utilisation du sol LiDAR et LANDSAT ont été successivement insérées dans le modèle en utilisant les mêmes paramètres de simulation. La durée de simulation a été réalisée sur 3 jours en utilisant les données météorologiques globales pendant la période du 1er janvier 2005 à 0h00 au 4 janvier à 24h00. Les résultats de ces simulations et leur comparaison sont décrits dans la partie suivante « Résultats ».

Résultats

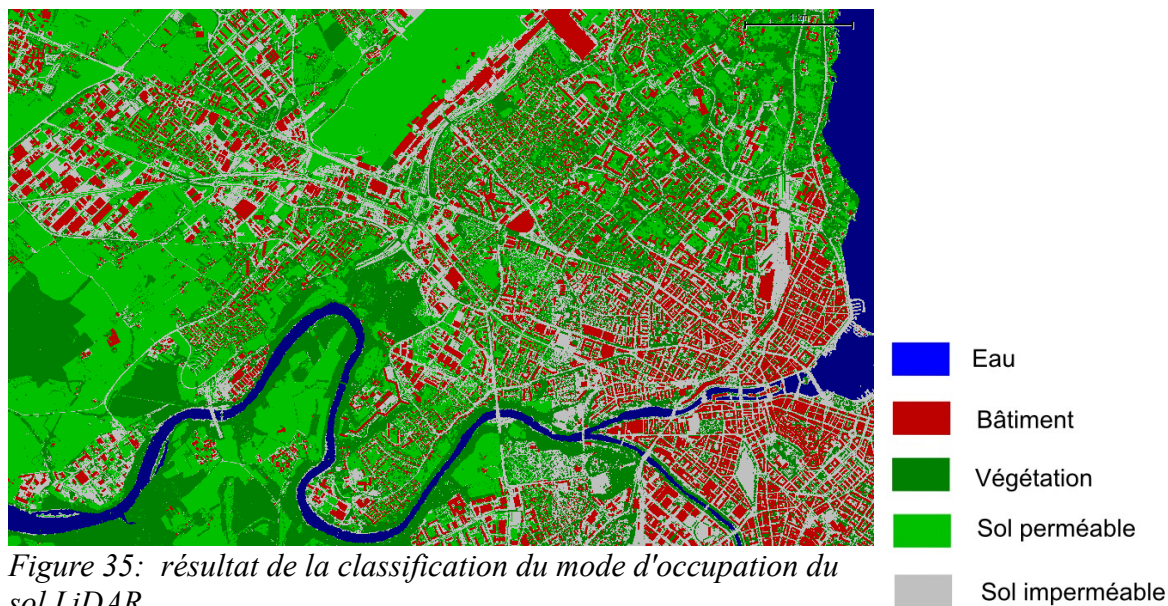
Cette partie expose les résultats des classifications issues des données LiDAR et des images LANDSAT ainsi que les résultats produits par le modèle FVM.

a. Résultats LiDAR

Dans un premier temps, le résultat de la classification LiDAR est présenté. Dans un second temps, le résultat produit par le modèle FVM et qui a utilisé les données d'utilisation du sol calculées à partir de cette classification est exposé.

Résultats des classifications LiDAR

Les résultats des classification réalisées avec les données LiDAR et la carte 2D du bâti sont présentés sur les deux figures suivantes. La figure 35 montre le résultat de la classification du mode d'occupation du sol issues des données LiDAR.



La figure 36 présente le résultat de la classification du mode d'utilisation du sol LiDAR. Elle a été produite à partir de la classification du mode d'occupation du sol issue des données LiDAR. Elle a été conçue pour obtenir les mêmes classes que celles présentes dans la classification LANDSAT et ainsi permettre une comparaison entre les données d'utilisation du sol LiDAR et les données d'utilisation du sol LANDSAT. C'est sur cette classification que les données d'utilisation du sol LiDAR ont été calculées.

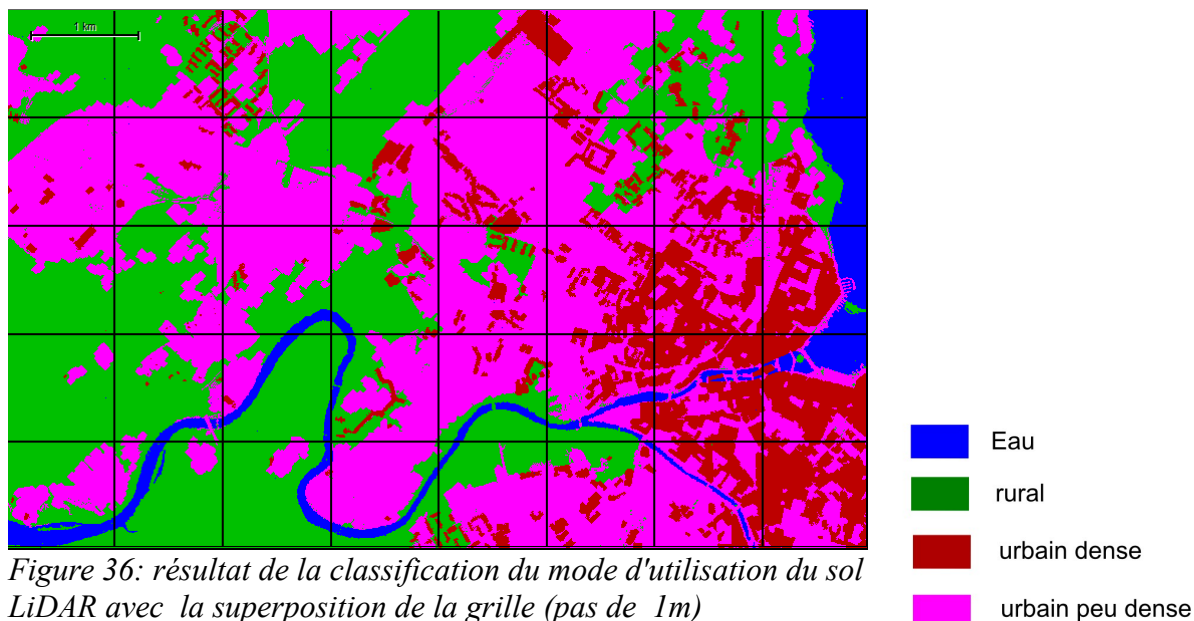


Figure 36: résultat de la classification du mode d'utilisation du sol LiDAR avec la superposition de la grille (pas de 1m)

Résultats LiDAR produits par le modèle FVM

Une première simulation a été réalisée en utilisant les données d'utilisation du sol calculées à partir de la classification d'utilisation du sol issue des données LiDAR. La durée de simulation a été réalisée sur 3 jours en utilisant les données météorologiques globales pendant la période du 1er janvier 2005 à 0h00 au 4 janvier à 24h00. La carte météorologique ci-dessous présente un résultat de cette simulation pour le 1er janvier 2005 à 21h sur une zone de 42km*42km centrée sur Genève.

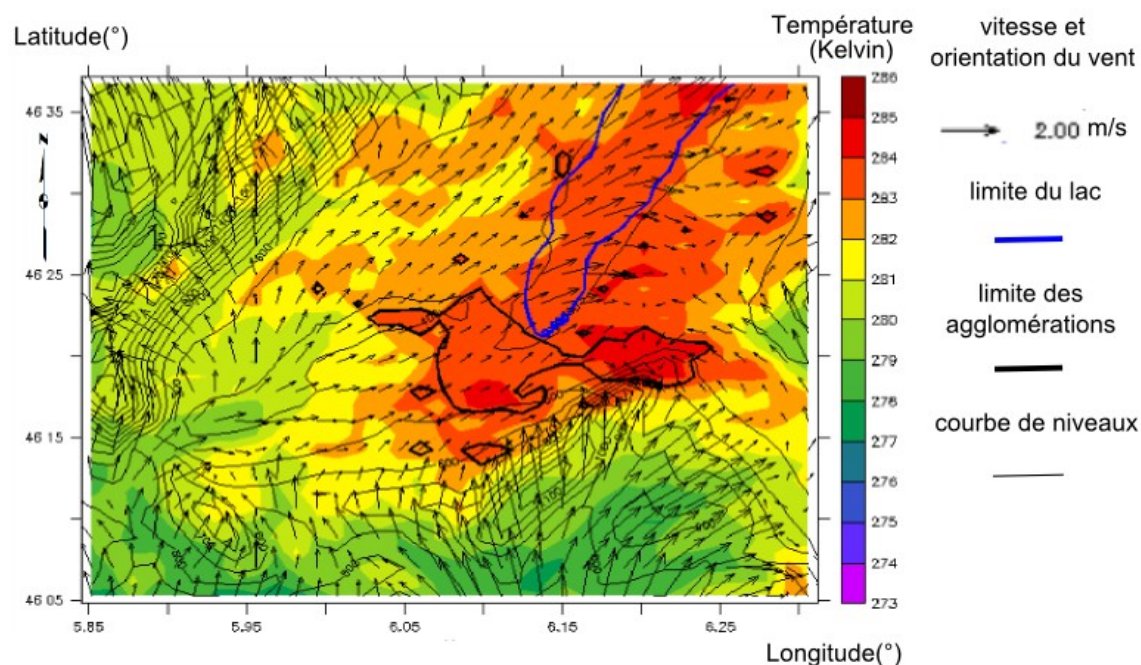


Figure 37: résultat de la simulation (1er janvier 2005 à 21h) avec les données d'utilisation du sol LiDAR

Lors de cette simulation à 21h le 1er janvier 2005 au niveau du sol, la ville est plus chaude que l'extérieur en raison de l'îlot de chaleur. Il en est de même pour le lac avec l'inertie thermique où l'eau se refroidit moins rapidement que le sol.

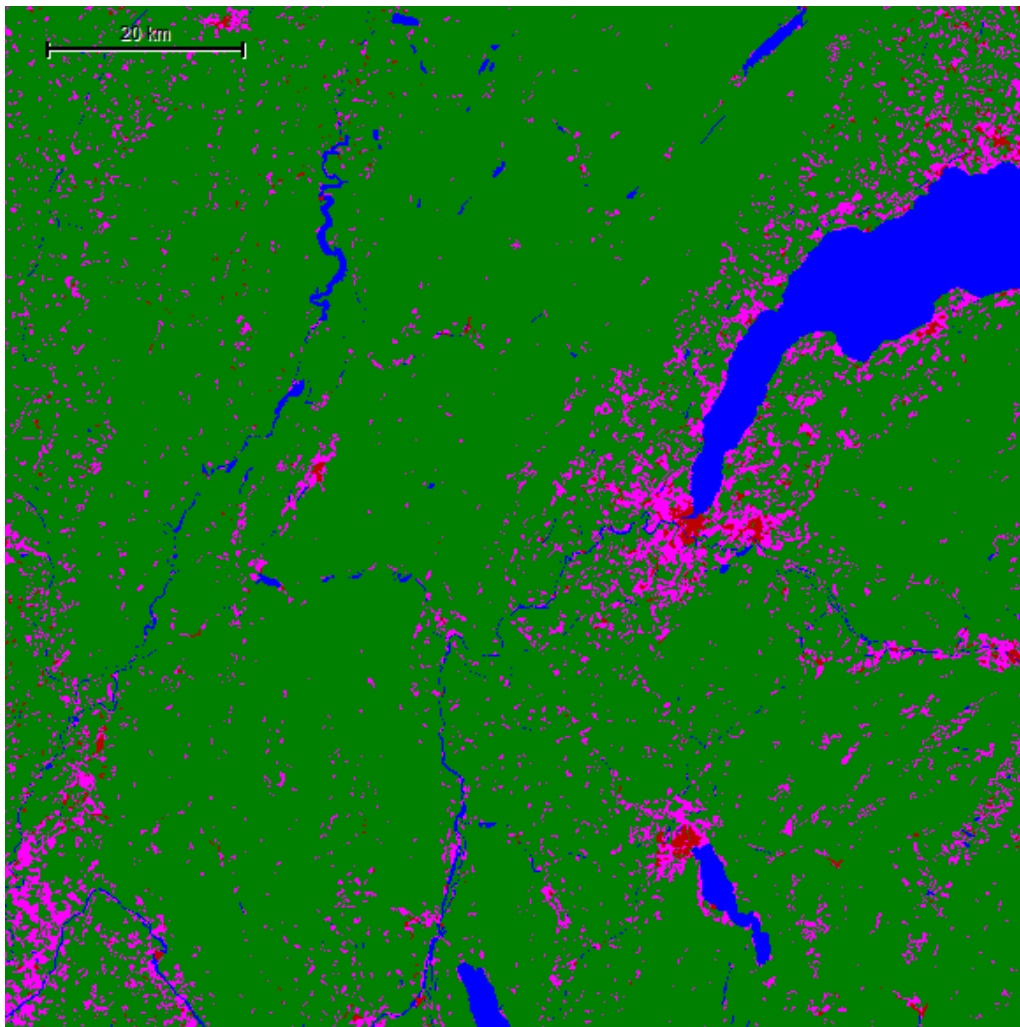
Les vents accélèrent au passage des reliefs au sud-est (Salève) et au nord-ouest (Jura). Dans le bassin genevois le vent est plus faible avec la protection des reliefs alentours. Il est canalisé et souffle du sud-ouest au nord-est. Emporté par le vent, un panache de température se forme à partir des zones les plus chaudes (ville et lac) vers le nord-est de la ville de Genève.

b. Résultats LANDSAT

Cette partie commence par présenter le résultat de la classification LANDSAT puis les résultats produits par le modèle FVM à partir des données d'utilisation du sol LANDSAT.

Résultats de la classification LANDSAT

Le résultat de la classification réalisée sur les images LANDSAT est présenté ci-dessous. C'est à partir de cette classification que les données d'utilisation du sol LANDSAT ont été calculées.



*Figure 38: résultat de la classification d'utilisation du sol LANDSAT 7 ETM+ (taille de l'empreinte : 105km*105km)*



La figure 39 présente le résultat de la classification du mode d'utilisation du sol LANDSAT réduite sur la même zone d'étude que la classification d'utilisation du sol LiDAR (8km*5km).

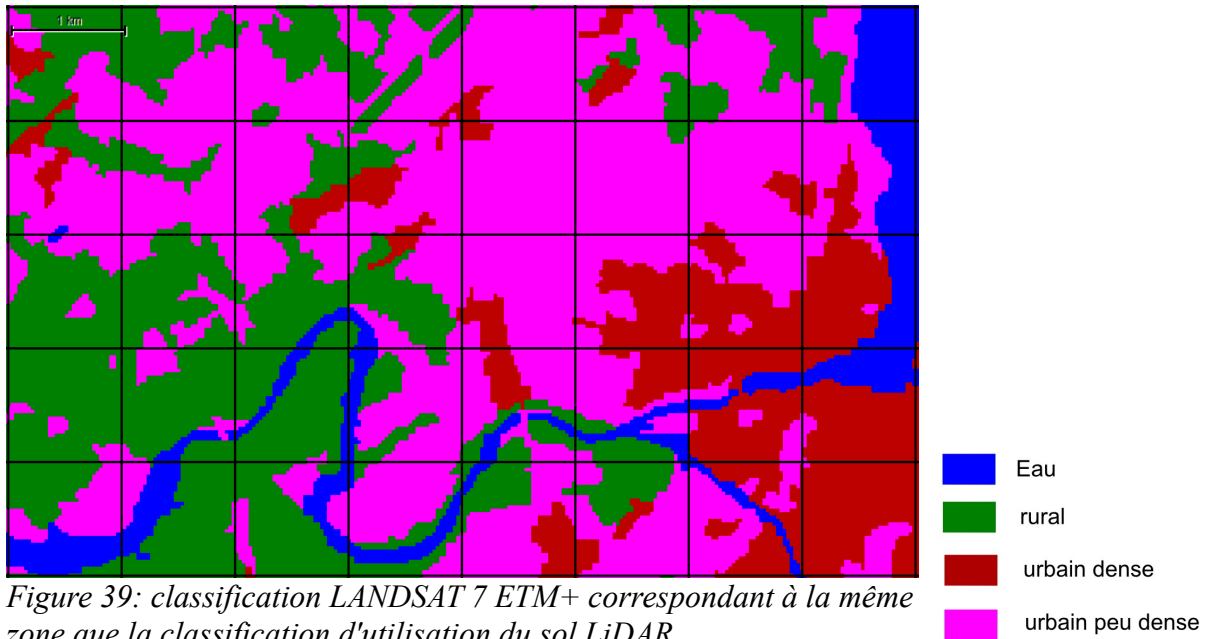


Figure 39: classification LANDSAT 7 ETM+ correspondant à la même zone que la classification d'utilisation du sol LiDAR

Résultats LANDSAT produits par le modèle FVM

Une seconde simulation a été réalisée en utilisant les données d'utilisation du sol calculées à partir de la classification issue des images LANDSAT. Les paramètres de cette simulation sont identiques à ceux de la première simulation basée sur les données d'utilisation du sol LiDAR, c'est à dire que la durée de simulation a été réalisée sur 3 jours en utilisant les données météorologiques globales pendant la période du 1er janvier 2005 à 0h00 au 4 janvier à 24h00. La carte météorologique ci-dessous présente un résultat de la simulation pour le 1er janvier 2005 à 21h. Elle représente la température au niveau du sol et les caractéristiques de vents (orientation et vitesse) sur la période et la zone considérée de la simulation.

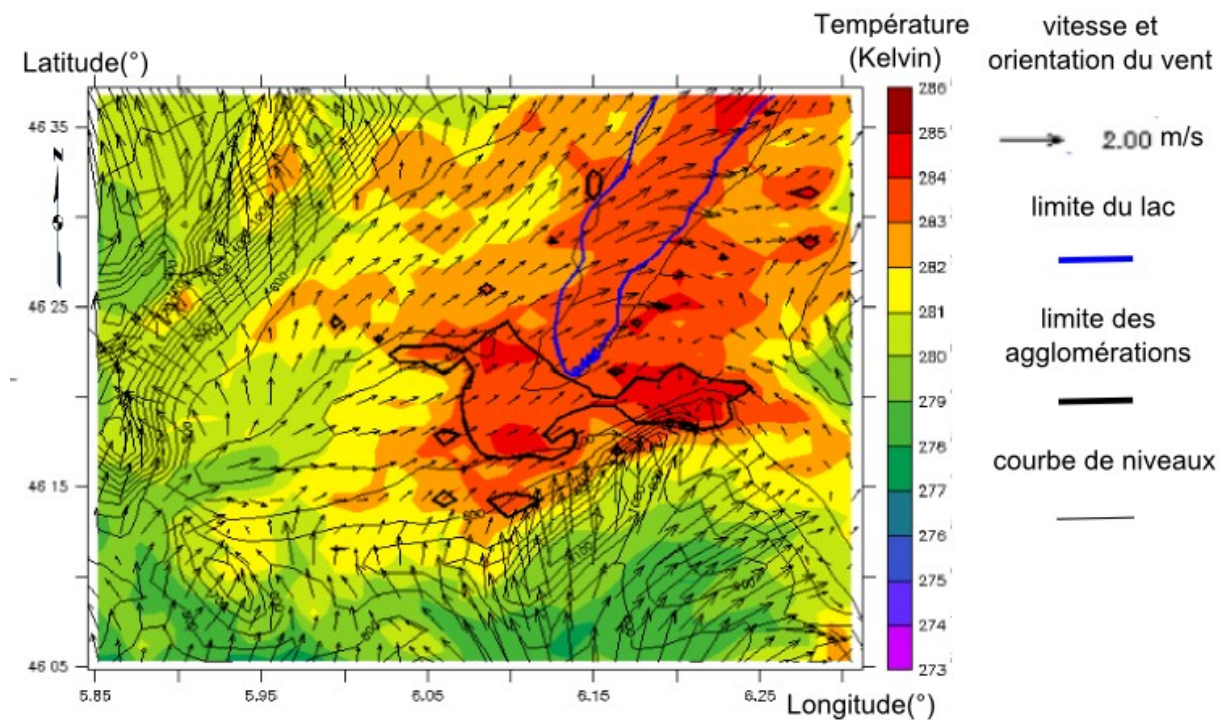


Figure 40: résultat de la simulation (1er janvier 2005 à 21h) avec les données d'utilisation du sol LANDSAT

L'interprétation des résultats de température et de vents sont globalement les mêmes que pour la simulation avec les données LiDAR avec quelques différences au niveau de la température qui vont être traitée dans la partie suivante.

c. Comparaisons LiDAR-LANDSAT

Pour constater une potentielle variation des résultats obtenus entre la simulation basée sur les données d'utilisation du sol issues des images LANDSAT et la simulation s'appuyant sur les données d'utilisation du sol issues des données LiDAR, une comparaison est réalisée entre ces deux simulations. Puis pour expliquer ces différences éventuelles dans la partie « Discussion », une comparaison effectuée directement sur les données d'utilisations du sol calculées à partir de la classification du mode d'utilisation du sol LiDAR et LANDSAT est présentée dans cette partie.

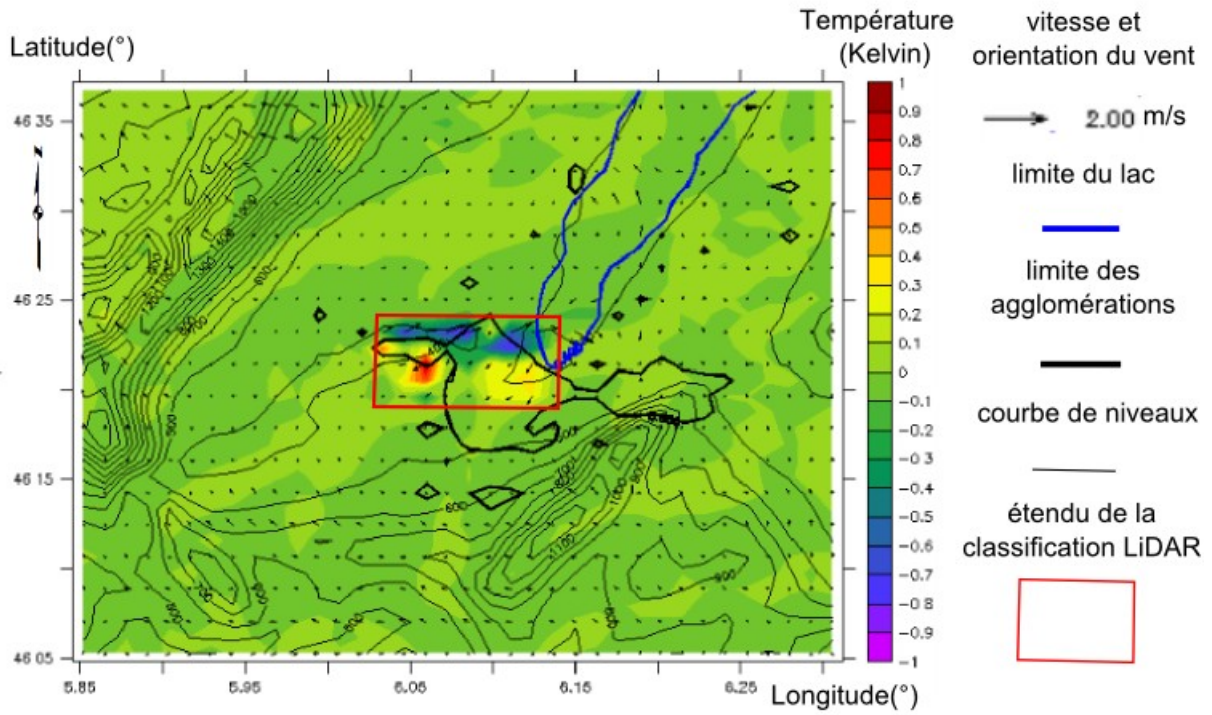


Figure 41: variation de température potentielle entre la simulation utilisant les données d'utilisation du sol LANDSAT et celle utilisant les données LiDAR

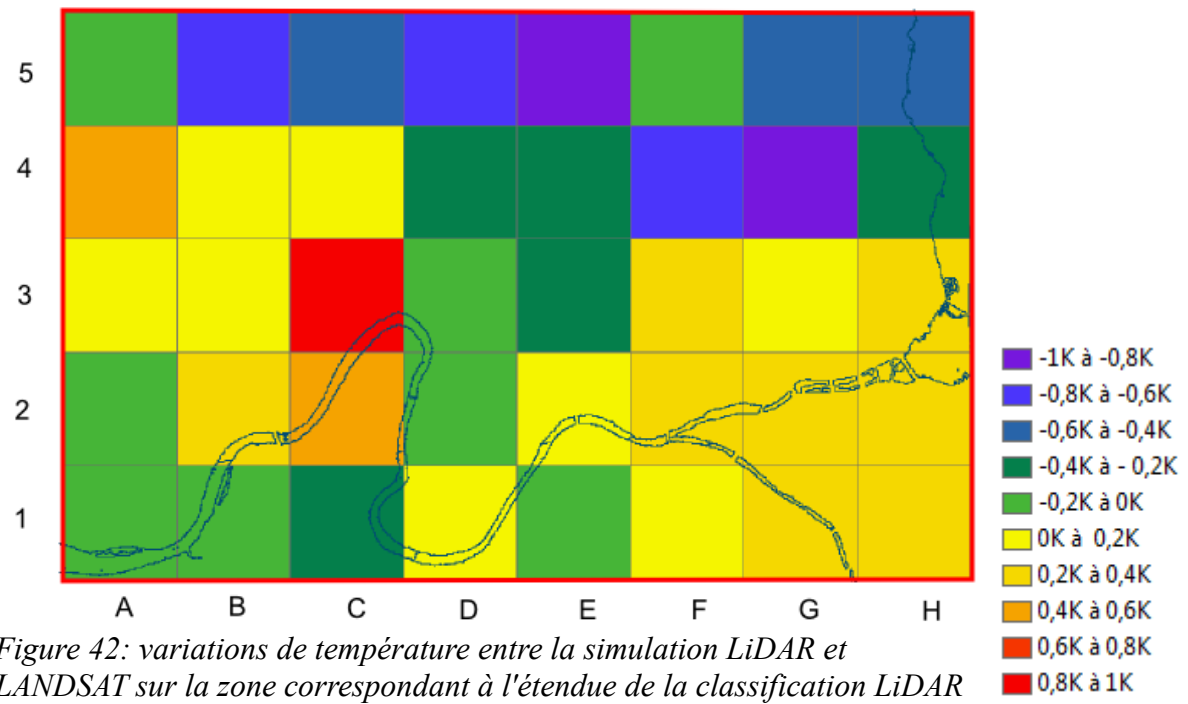


Figure 42: variations de température entre la simulation LiDAR et LANDSAT sur la zone correspondant à l'étendue de la classification LiDAR

Discussion

Les variations de températures entre les deux simulations sont inférieures à 1K. Il semblerait que le modèle puisse mettre en évidence des différences entre les deux bases de données d'occupation du sol.

Les différences entre les deux simulations se situent au niveau de la ville, à l'intérieur de l'étendue de la classification LiDAR. Ceci s'explique logiquement par le fait que les « données d'utilisation du sol LiDAR » sont prises égales aux « données d'utilisation du sol LANDSAT » à l'exception de la zone commune aux deux classifications. En effet, pour concevoir les « données d'utilisation du sol LiDAR », les « données d'utilisation du sol LANDSAT » ont été modifiées sur l'étendue de la « classification LiDAR » pour les remplacer par les valeurs calculées à partir de la classification du mode d'utilisation du sol basée sur les données LiDAR. Dans la suite de cette partie, l'analyse des différences constatées se limitera à la zone correspondant à l'étendue de la classification LiDAR (Cf figures 41 et 42)..

Un des écarts les plus importants se trouve au niveau de la cellule C3 de la figure 42 avec une variation de +0,85K. En comparant les données d'utilisation du sol dans cette cellule, on constate que les "données d'utilisation du sol LiDAR" sont constituées de +28% d'urbain peu dense et de -28% de rural par rapport aux « données d'utilisation du sol LANDSAT ». La classification issue des données LiDAR a détectée une part plus importante de zone urbaine peu dense au détriment de la zone rurale. Cela s'explique par le fait que cette cellule est constituée d'une zone pavillonnaire très végétalisée avec une présence importante de jardins, arbres et arbustes. La classification réalisée sur l'image LANDSAT a donc classée cette zone résidentielle comme une classe « rural » car sa résolution de 30m n'est pas assez fine pour obtenir une segmentation et des critères de classification suffisants pour détecter les maisons individuelles dans ce cas là. Cette classification de la zone résidentielle en classe « rural » explique pourquoi il y a un pourcentage de données d'utilisation du sol appartenant à la classe « urbaine peu dense » plus faible dans le cas de l'image LANDSAT. Le modèle FVM interprète cette différence en considérant que la température est plus élevée avec les données d'utilisation du sol LiDAR puisque la cellule est constituée d'une proportion plus importante de zone urbaine qui va créer un effet d'îlot de chaleur supplémentaire au niveau de cette cellule.

Au niveau du centre ville de Genève (carré formé par les cellules G1, H1, G2 et H2), les écarts constatés par le modèle FVM sont en moyenne de 0,3K. Dans ces cellules, « les « données d'utilisation du sol LiDAR » sont constituées en moyenne de +18% de surface « urbain peu dense », -15% de « urbain dense » et de -3% de surface « eau ». La classification basée sur les données LiDAR a détecté une part de zone « urbain peu dense » au détriment d'une part importante de surface « urbain dense » et d'une part de surface « eau ». Cette différence s'explique par l'utilisation d'une méthode de classification plus restrictive des surfaces « dense » avec la mise en place de critères liés à la hauteur des bâtiments et la distance entre les bâtiments dans le cadre de la classification « LiDAR ». De plus, dans la classification du mode d'utilisation du sol issue des images LANDSAT, la segmentation de l'image a créé des objets avec des formes simplifiées et une surface plus importante, ce qui représente les cours d'eau de manière plus large et épais que dans la classification « LiDAR » (Cf figure 36 et 39). Les surfaces « eau » sont donc plus importantes dans la « classification LANDSAT ». Le modèle FVM interprète cette différence en considérant que la température est plus élevée avec les données d'utilisation du sol LiDAR puisque la cellule est constituée d'une proportion moins importante de surface

« eau », ce qui accentue légèrement le phénomène d'îlot de chaleur urbain et contribue à augmenter la température dans le cas de la « simulation LiDAR ». En outre, on peut remarquer que l'influence des surfaces « urbain dense » et « urbain peu dense » sur la température de l'air sont équivalents. En effet, la différence de ces données d'utilisation du sol entre les deux classifications est importante (+15%). On aurait pu penser que dans la « classification LANDSAT » où la part de la surface « urbain dense » est plus importante (15%) que dans « la classification LiDAR », l'îlot de chaleur urbain allait être plus marqué et allait ainsi contribuer à augmenter la température. Or, c'est l'inverse qui s'est produit. On peut conclure que la variation des effets des surfaces « urbaines dense » et « urbain peu dense » est faible. Par contre, une faible variation de la surface « eau » (3%) va avoir plus d'influence sur les résultats produits par le modèle FVM qu'une variation élevée des surfaces « urbain dense » et « urbain peu dense » (15%). Il est donc important d'avoir une classification détaillée des cours d'eau et des lacs et la distinction des surfaces « urbain dense » et « urbain peu dense » n'a pas besoin d'être affinée.

Un autre écart ayant pour valeur -0,87K se trouve au niveau de la cellule E5. Dans cette cellule, les « données d'utilisation du sol LiDAR » sont constituées de +30% de surface « rurale », de -35% de surface « peu dense » et de +5% de surface « dense ». La classification issue des données LiDAR a détecté des parts plus importantes de zone « rural » à la place d'une zone d'« urbain peu dense ». Cette différence s'explique en raison du modèle numérique d'intensité utilisée dans la « classification du mode d'occupation LiDAR ». En effet, dans le modèle d'intensité, les routes et les éléments asphaltés ou bitumés sont plus sombres que les éléments liés au sol perméable tel que les prairies ou le sol nu. Il a donc été considéré que les objets qui ont une intensité inférieure à la valeur 30 appartiennent à la classe « sol imperméable » pour les distinguer du sol « perméable ». Or dans la cellule E5, un cas particulier s'est présenté avec la piste d'aéroport composée de macadam qui a un albédo plus élevé que les surfaces asphaltés, bitumés ou avec du ciment. Par conséquent, la piste d'aéroport n'a pas été détectée comme « sol imperméable » dans la méthode de la classification du mode d'occupation du sol issue des données LiDAR. Pour cette cellule, la classification du mode d'utilisation du sol ne doit pas être considérée comme ayant une meilleure précision que les images LANDSAT. Cette explication s'applique également pour la cellule G4 où il y a également une surface « rural » plus importante dans la « classification LiDAR » que dans la « classification LANDSAT ».

En comparant visuellement les deux classifications des modes d'utilisation du sol (Cf figure 36 et 39), on observe que les limites des modes d'utilisation du sol dans le cas de la « classification LiDAR » sont plus nettes que pour la « classification LANDSAT » où les limites sont plus dentelées. Même si la forme n'a pas d'importance pour le modèle, cette différence de représentation des limites crée une variation des surfaces de chacune des classes dans les cellules. Par exemple, dans la classification du mode d'utilisation du sol issue des images LANDSAT, la segmentation de l'image a créé des objets avec des formes simplifiées et une surface plus importante. On constate visuellement que les cours d'eau sont plus larges que dans la classification « LiDAR ». Les surfaces de la classe « eau » ont été calculées sur ces deux classifications et confirme ce qui a été constaté visuellement. La superficie de la classe « eau » représente une superficie de 6% de la superficie totale de la « classification LiDAR » contre 8% avec la classification « LANDSAT ». De plus, dans la classification du mode d'utilisation du sol issue des données LiDAR, les objets classés « urbain denses » sont plus fragmentés que les autres

Conclusion

Le LiDAR contribue à améliorer les résultats produits par le modèle FVM. Il permet de classier de manière plus précise les zones urbaines peu denses avec la détection des zones résidentielles et permet d'affiner la classification de l'eau. Mais cette amélioration reste faible compte tenu des écarts constatés (différence de températures inférieure à 0,9K).

Le LiDAR doit être combiné avec l'utilisation d'autre sources de données pour devenir un apport pour le modèle FVM. Cette combinaison de données permet de faciliter et améliorer la classification adaptée au modèle FVM. La carte 2D du bâti permet de distinguer la végétation des bâtiments. Le modèle numérique d'intensité interpolé à partir du nuage de points LiDAR fournit des informations intéressantes pour classier des modes d'occupation du sol (sol perméable et sol imperméable). Mais il aurait fallu également utiliser d'autre sources de données comme des orthophotos pour affiner la classification.

Références bibliographiques

AXELSSON P., *Processing of laser scanner data - algorithms and applications*, 1999 The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 54, Part 2, pp. 138-147.

BRENNAN R., WEBSTER T.L., *Object-oriented land cover classification of lidar-derived surfaces*, Can. J. Remote Sensing, Vol. 32, No. 2, pp. 162–172, 2006
Disponible sur http://agrg.cogs.nsc.ca/gallery2/main.php?g2_view=core.DownloadItem&g2_itemId=815

CARNEIRO C., *Extraction of urban environmental quality indicators using LiDAR-based Digital Surface Models*, 2011, thèse de l'École Polytechnique Fédérale de Lausanne

GONÇALVES G., *Analysis of interpolation errors in urban digital surface models created from LIDAR data*, 2006, Proceedings of the 7th International Symposium on Spatial Accuracy, Assessment in Resources and Environment Sciences. pp. 160-168.

HAFTENDORN E., *Classification de l'utilisation du sol de la ville de Lahore et de ses environs dans le but de modéliser l'îlot de chaleur urbain*, 2010, projet Sciences et Ingénierie de l'Environnement de l'École Polytechnique Fédérale de Lausanne

KRPO A., *Development and Application of a Numerical Simulation System to Evaluate the Impact of Anthropogenic Heat Fluxes on Urban Boundary Layer Climate*, 2008, thèse de l'École Polytechnique Fédérale de Lausanne
Disponible sur http://biblion.epfl.ch/EPFL/theses/2009/4428/EPFL_TH4428.pdf

LUXORION, *Les effets locaux du vent*, Site disponible sur <http://www.astrosurf.com/luxorion/meteo-vent2.htm> (Page consultée le 01/05/11)

MALLET C., *Analyse de données lidar à Retour d'Onde Complète pour la classification en milieu urbain*, 2010, Thèse de Télécom ParisTech
Disponible sur http://recherche.ign.fr/labos/matis/pdf/theses/mallet_thesis_web.pdf

MARTILLI A., *Development of an urban turbulence parameterisation for mesoscale atmospheric models*, 2001, thèse de l'École Polytechnique Fédérale de Lausanne
Disponible sur http://lpas.epfl.ch/MOD/publi/PhD/PhD_martilli.pdf

ROULET Y.-A., *validation and application of an urban turbulence parameterisation scheme for mesoscale atmospheric models*, 2004, thèse de l'École Polytechnique Fédérale de Lausanne
Disponible sur http://biblion.epfl.ch/EPFL/theses/2004/3032/EPFL_TH3032.pdf

TARSHA KURDI F., *Extraction et reconstruction de bâtiments en 3D à partir de relevés lidar aéroportés*, 2008, Thèse de l'Université Louis Pasteur Strasbourg I
Disponible sur http://scd-theses.u-strasbg.fr/1528/01/TARSHA-KURDI_Fayez_2008.pdf

Site du groupe pollution de l'air et climat urbain du laboratoire Image, Ville, Environnement de Strasbourg : <http://imaville.u-strasbg.fr/spip.php?rubrique23>

"Modelisation dynamique et thermodynamique de la canopee urbaine : realisation du modele de sols urbains pour submeso", DUPONT Sylvain, 2001, Thèse de l'université de Nantes

http://tel.archives-ouvertes.fr/docs/00/44/37/56/PDF/these_sylvain_dupont.pdf

"La ville : climat, pollutions et végétal", Yves Richard, http://www.geographiedijon.fr/IMG/pdf/L3_pro_GPPVMU_1.pdf

"Impact des modifications de l'albédo et des matériaux urbains sur la température et la qualité de l'air à Madrid", Tristan Mariethoz, 2006, projet de master SSIE – ENAC – EPFL

"Building model reconstruction from lidar data and aerial photographs", Ruijin Ma, 2004, Thèse de "The Ohio State University"

<http://etd.ohiolink.edu/send-pdf.cgi/Ma%20Ruijin.pdf?osu1104114425>

"terrain classification using airborne lidar data and aerial imagery", O. Brattberg*, G. Tolt, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B3b. Beijing 2008*

http://www.isprs.org/proceedings/XXXVII/congress/3b_pdf/55.pdf

RASHEED A., *Multiscale Modelling of Urban Climate*, 2009, thèse de l'École Polytechnique Fédérale de Lausanne

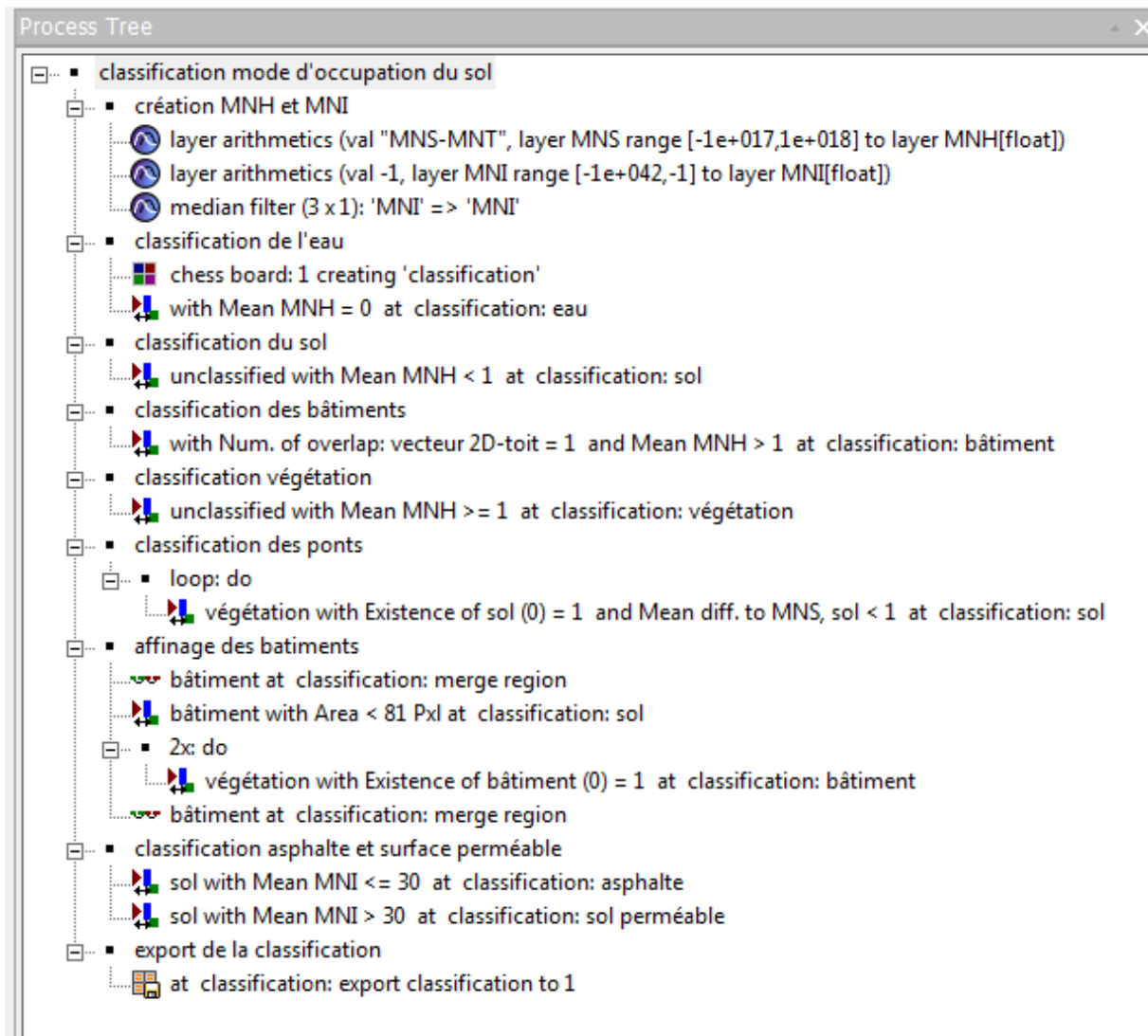
Disponible sur http://biblion.epfl.ch/EPFL/theses/2009/4531/EPFL_TH4531.pdf

"Résumé d'orientation des Directives de l'OMS pour la Qualité de l'Air", OMS, 2000

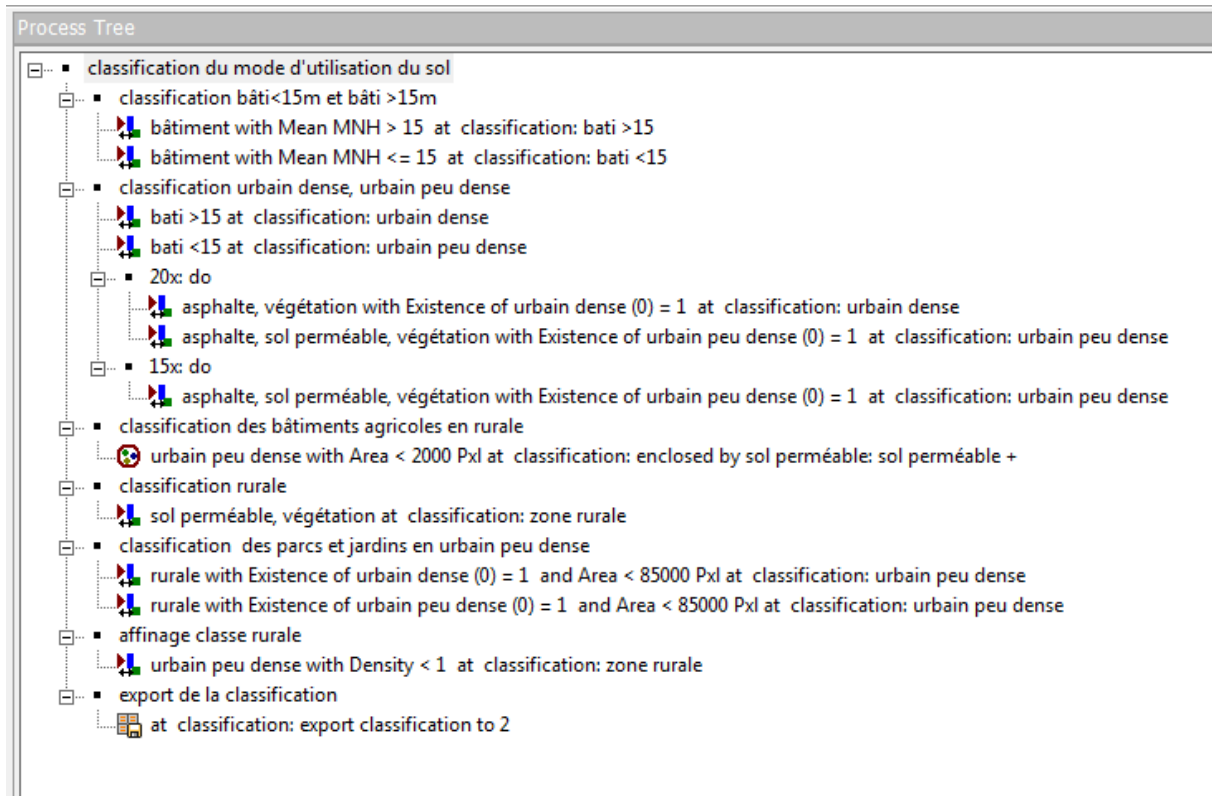
https://apps.who.int/environmental_information/Air/Guidelines/ResumeF.htm

Annexe

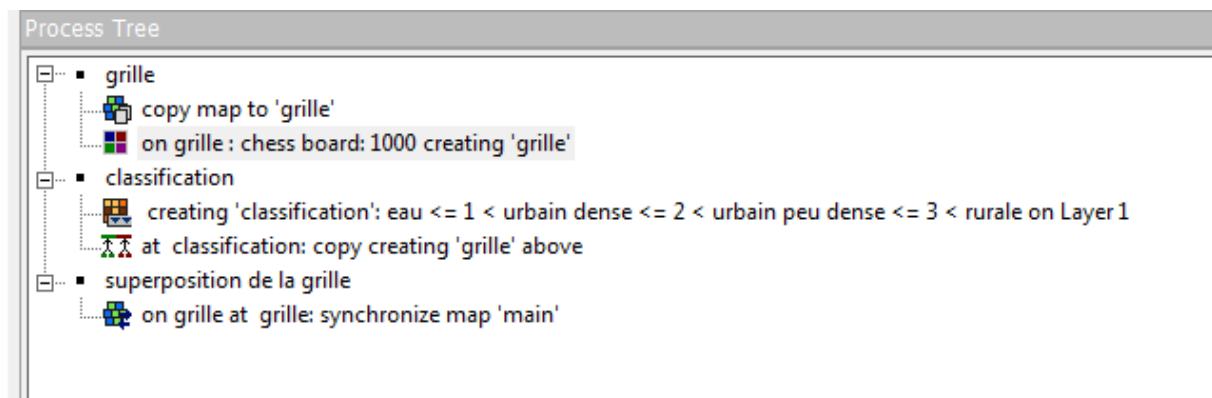
A. Chaîne d'algorithmes pour la classification du mode d'occupation du sol



B. Chaîne d'algorithmes pour la classification du mode d'utilisation du sol



C. Chaîne d'algorithmes pour la création de la grille et sa superposition sur la classification du mode d'utilisation du sol

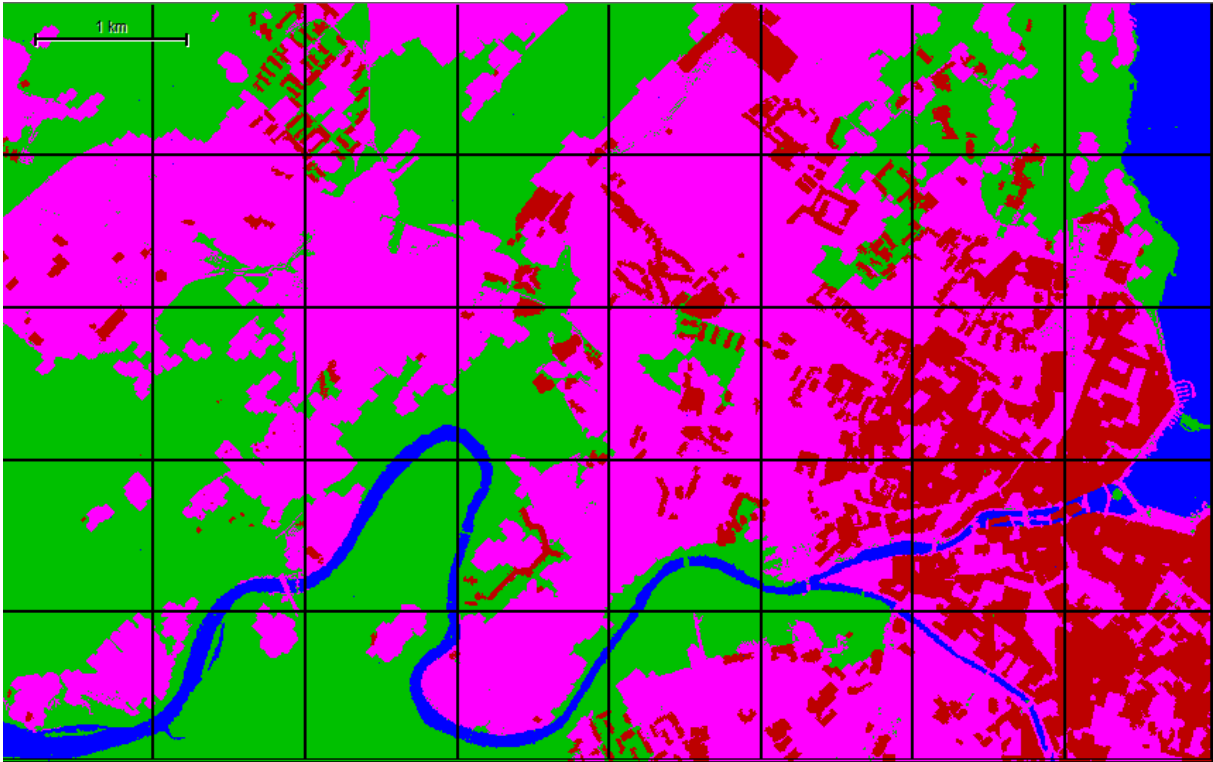


Le principe consiste à créer un plan (« map ») sur lequel la grille va être créée. Ensuite, dans un autre plan, la classification va être réalisée. Cette classification va être copiée au-dessus dans le même plan. Pour finir, la grille va être synchronisée sur la classification qui a été copiée.

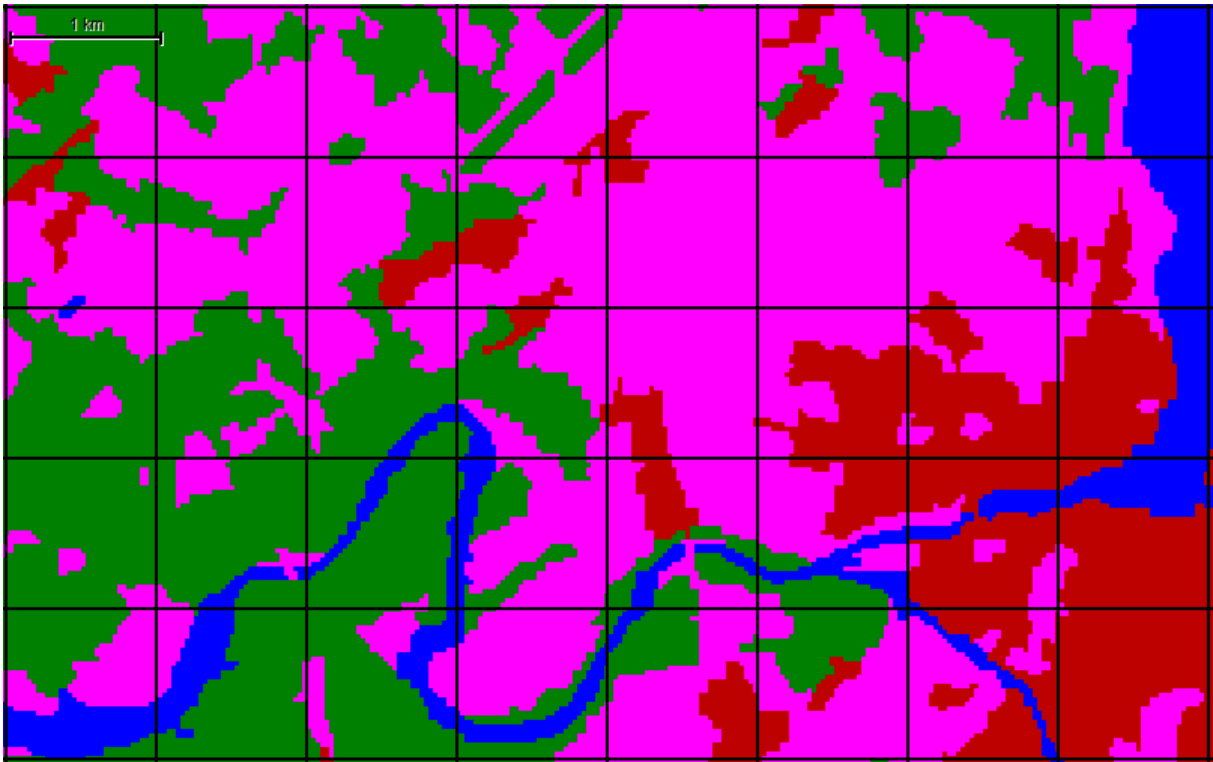
remarque : Dans l'algorithme « synchronize map main », le paramètre « synchronize complete hierarchy » ne doit pas être activé (no).

Pour obtenir les pourcentages surfaciques, le critère « Rel. Area of » a été utilisé.

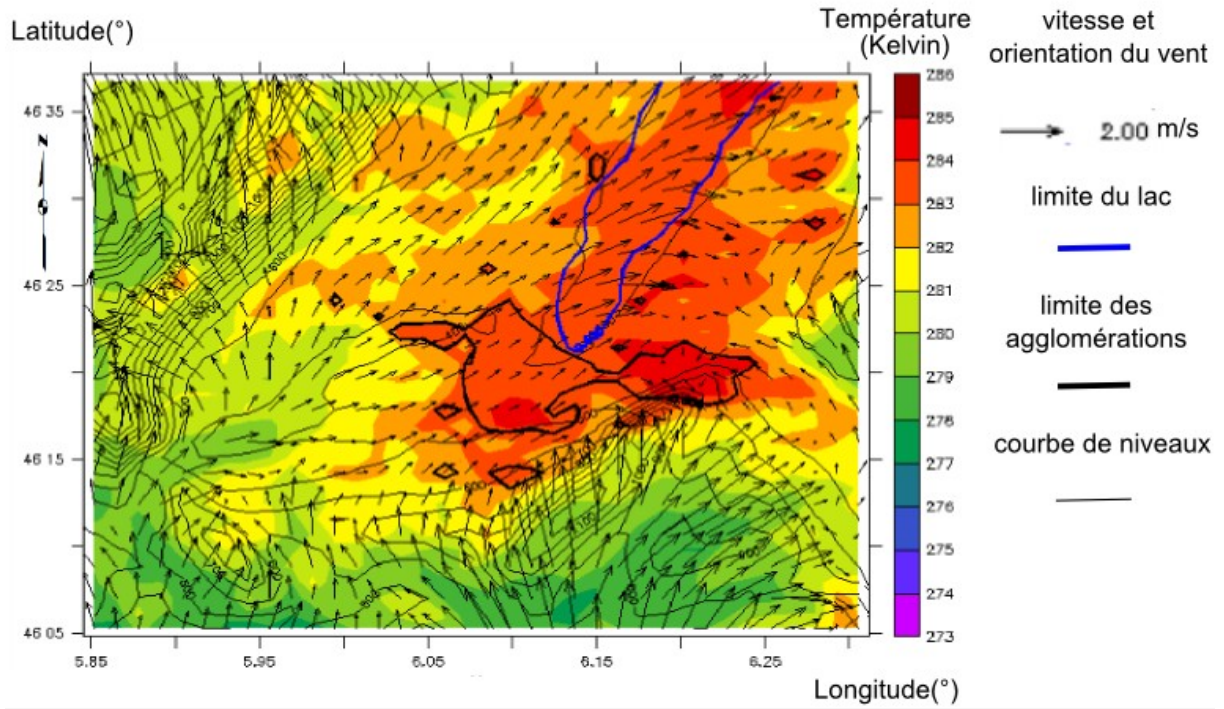
D. Classification du mode d'utilisation du sol LiDAR



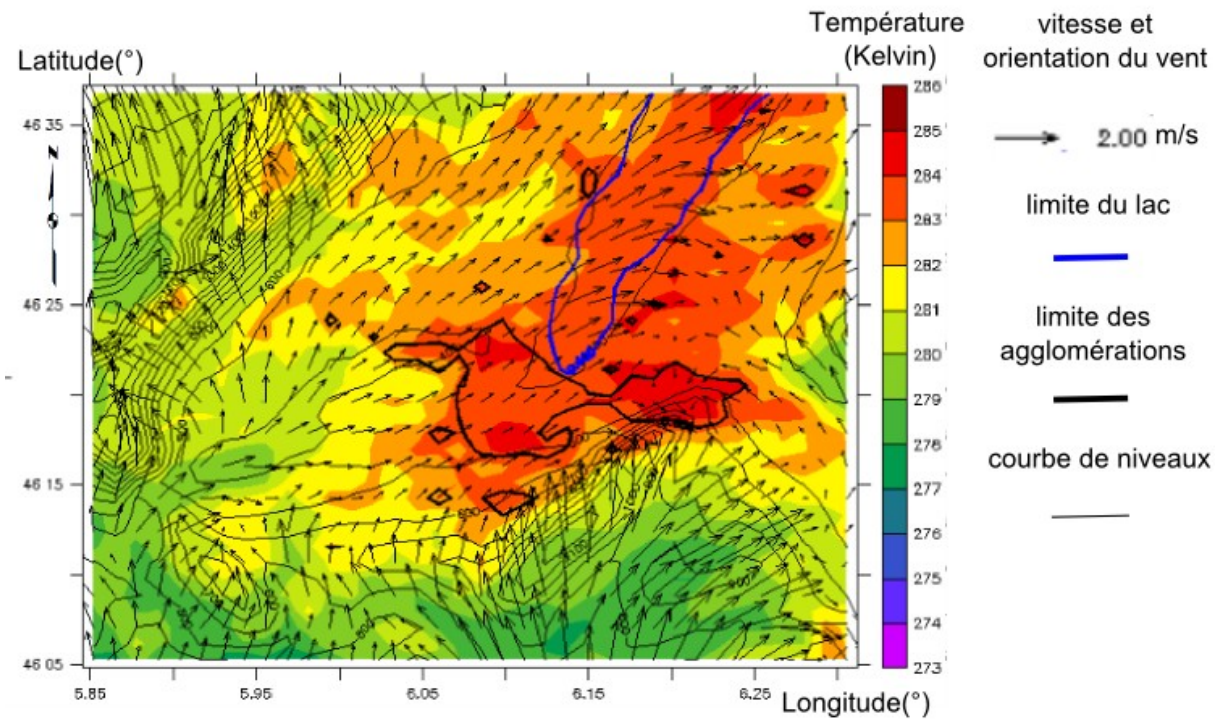
E. Classification du mode d'utilisation du sol LANDSAT



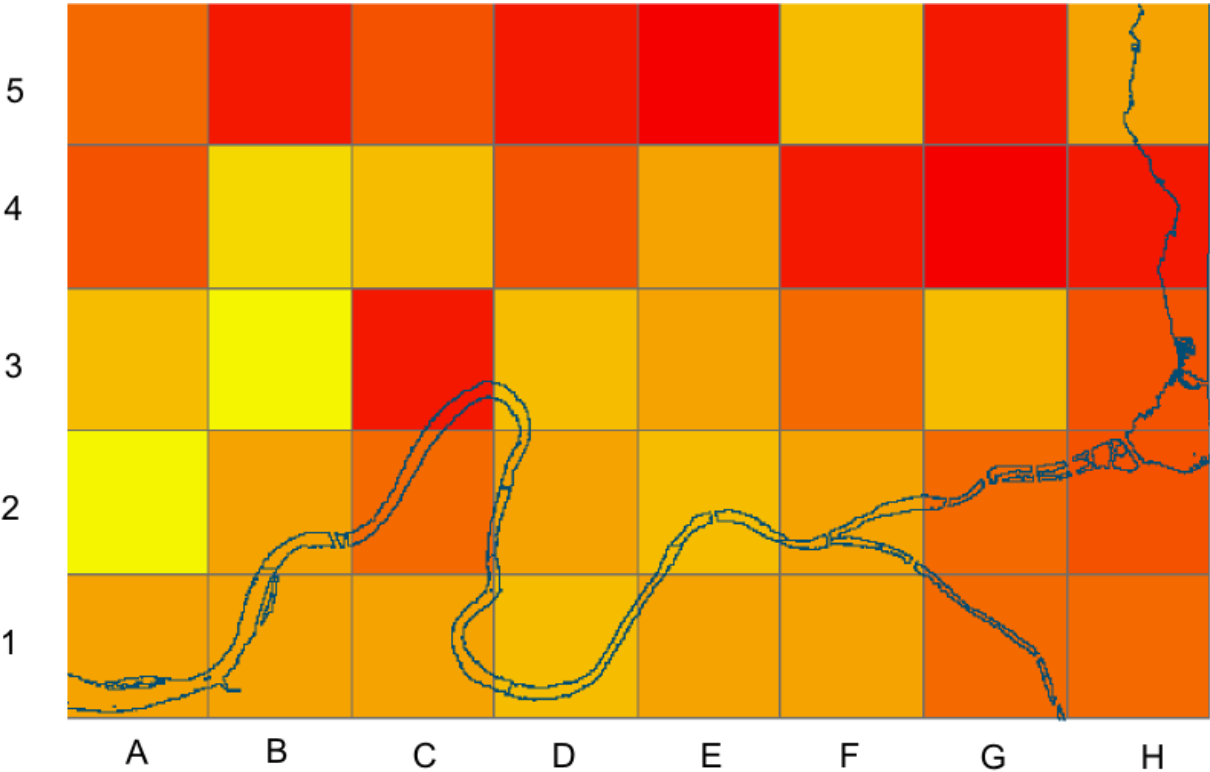
F. Résultat de la simulation LiDAR



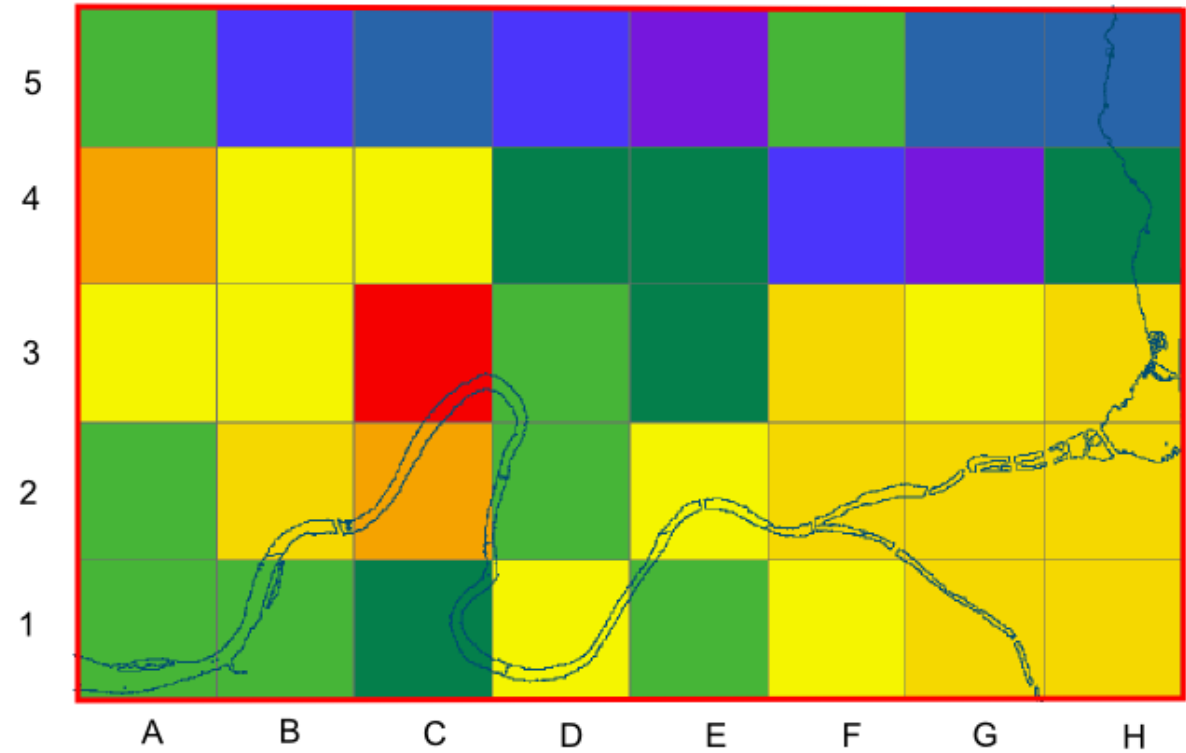
G. Résultat de la simulation LANDSAT



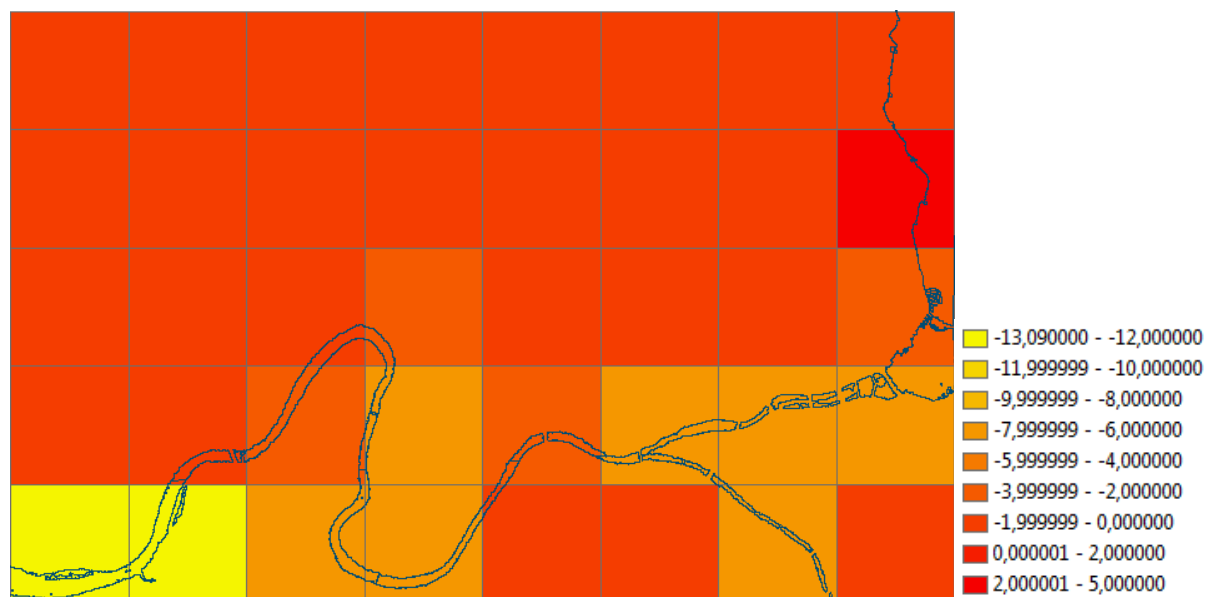
H. Comparaison des données d'utilisation du sol LiDAR et LANDSAT



I. Comparaison des simulations LiDAR et LANDSAT



J Comparaison des pourcentages surfacique de la classe eau (LiDAR-LANDSAT)



K Comparaison des classes sur la totalité de la scène entre la classification LiDAR et LANDSAT

	classification LiDAR	classification LANDSAT	
eau (surface)	2357400	3207000	
rural	13792000	11581700	
urbain dense	5565200	6385500	
urbain peu dense	18285400	18825800	
	40000000	40000000	
			ecart
Eau (% de la scene))	0,06	0,08	-0,02
rural	0,34	0,29	0,06
urbain dense	0,14	0,16	-0,02
urbain peu dense	0,46	0,47	-0,01