

Note sur l'expérimentation de dénombrement des arbres sur les communes de Lille, Lomme et Hellemmes à partir des données LiDAR

Ville de Lille - J. Benoist – Août 2022 – Révision Décembre 2022

Introduction

Liée à la transition et à diverses thématiques écologiques, la question des services rendus par les arbres est plus que jamais d'actualité. Or, la simple question de leur dénombrement est toujours posée, qu'il s'agisse de ceux gérés en propre par les villes, qui ne disposent pas toujours d'un inventaire complet, ou de ceux sur terrains privés, plus difficilement accessibles.

Pour obtenir des informations, le relevé de terrain est évidemment le plus précis, mais c'est également le plus complexe à mettre en œuvre, tant en termes d'équipement (GPS, tablettes...), de formation des agents, que de temps. Depuis quelques années, des sociétés de service spécialisées en traitement des images satellitaires proposent des données de plus en plus complexes de classification des sols et des espaces, de plus en plus complétées par des produits dérivés (qualité de l'air...). La mesure de la canopée fait partie des données qu'il est possible de commander.

Par ailleurs, depuis quelques années, la Métropole Européenne de Lille (MEL), comme d'autres collectivités, a entrepris de procéder à une reconnaissance de son territoire par la technologie LiDAR aéroportée et propose des outils simples de visualisation, l'accès aux données brutes. Elle effectue également des traitements en fonction des thématiques diverses des politiques urbaines (urbanisme, environnement...).

Pour rappel, le LiDAR permet d'obtenir un « nuage de points » virtuel, dont la densité permet de recréer un modèle tridimensionnel de l'environnement d'une grande précision. Cependant, le volume de données implique l'utilisation de nouveaux outils et de nouvelles méthodes encore peu répandues.

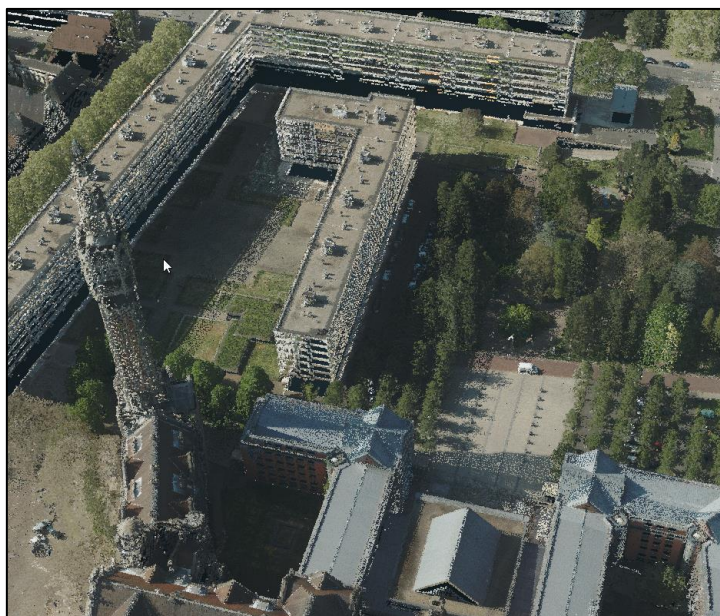


Figure 1 - Vue du beffroi de Lille avec l'outil en ligne Potree. 98 points par m2 permettent de reproduire fidèlement la réalité et même de visiter virtuellement le modèle.

Suivant de près cette technologie, j'ai proposé de réaliser une première étude en interne afin de déterminer s'il était possible d'obtenir très rapidement un chiffre à partir de ces données.

Choix de la méthode

Pour réaliser le décompte des arbres à l'aide d'un nuage de points, peu de solutions s'offrent à nous :

- Le comptage « manuel » à partir d'outils de visualisation et de « coupe » et classification semi-automatisée du nuage, très chronophage.
- Quelques outils professionnels dont le coût des licences est prohibitif (Terrasolid, ArcGIS Pro...) et qui ne proposent souvent par défaut qu'une classification standard.
- Des méthodes plus ou moins expérimentales et plus ou moins abouties, très en vogue notamment du fait du développement des voitures autonomes.

Après une assez longue recherche bibliographique et le test d'outils de traitement spécifique des données arborées (LASTools, Whitebox tools, FUSION...), mon choix s'est orienté vers les travaux de Jean-Romain Roussel, chercheur à l'Université Laval au Québec.

« Quantification théorique des effets du paramétrage du système d'acquisition sur les variables descriptives du nuage de points LiDAR. » - J.R. Roussel 2018.

En effet, celui-ci a développé une suite d'outils (1) qui est aujourd'hui reconnue comme une référence, nommée lidR. Poursuivant son travail avec le soutien du Ministère des Forêts, de la Faune et des Parcs, de l'Université Laval (Québec) et de l'Université de Colombie Britannique (Canada), il en a étendu les capacités en implémentant différents algorithmes issus des plus récentes recherches scientifiques. D'autres packages entrent en jeu grâce à la philosophie du code source ouvert.



lidR, une suite d'outils spécialisés

En quelques mots, il est possible de créer une chaîne de traitements automatisés qui va répondre à tous les besoins spécifiques de la gestion des données forestières à partir de données LiDAR aéroportées. Voici quelques points forts de lidR :

- Développé en R, un langage spécialisé en traitements statistiques, ce qui permet de profiter de nombreux outils existants ou en cours de développement, tels les packages raster, terra, sf, EBIImage ...
- Code source ouvert ;
- Support total des formats usuels, notamment les tuiles mises à disposition par la MEL avec gestion automatique des recouvrements, des images TIFF géo-localisées, des géo-packages ;
- Outils nombreux et extensibles de traitement des données ;
- Localisation automatique des arbres, segmentation, création de modèles numériques de terrain, d'élévation, de canopée, traitements statistiques ;
- Rapidité d'exécution grâce à l'utilisation du langage C pour certains outils ;
- Une documentation complète et facile d'accès, ainsi qu'un support technique.

Détail de la chaîne de traitements expérimentale

- 1- Installation de R, du package lidR et des autres packages nécessaires à son utilisation (en mode utilisateur ou sur plateforme web)

- 2- Téléchargement des tuiles LiDAR de l'emprise du projet (éventuellement utiliser la couche tindex.gpkg pour automatiser cette tâche).

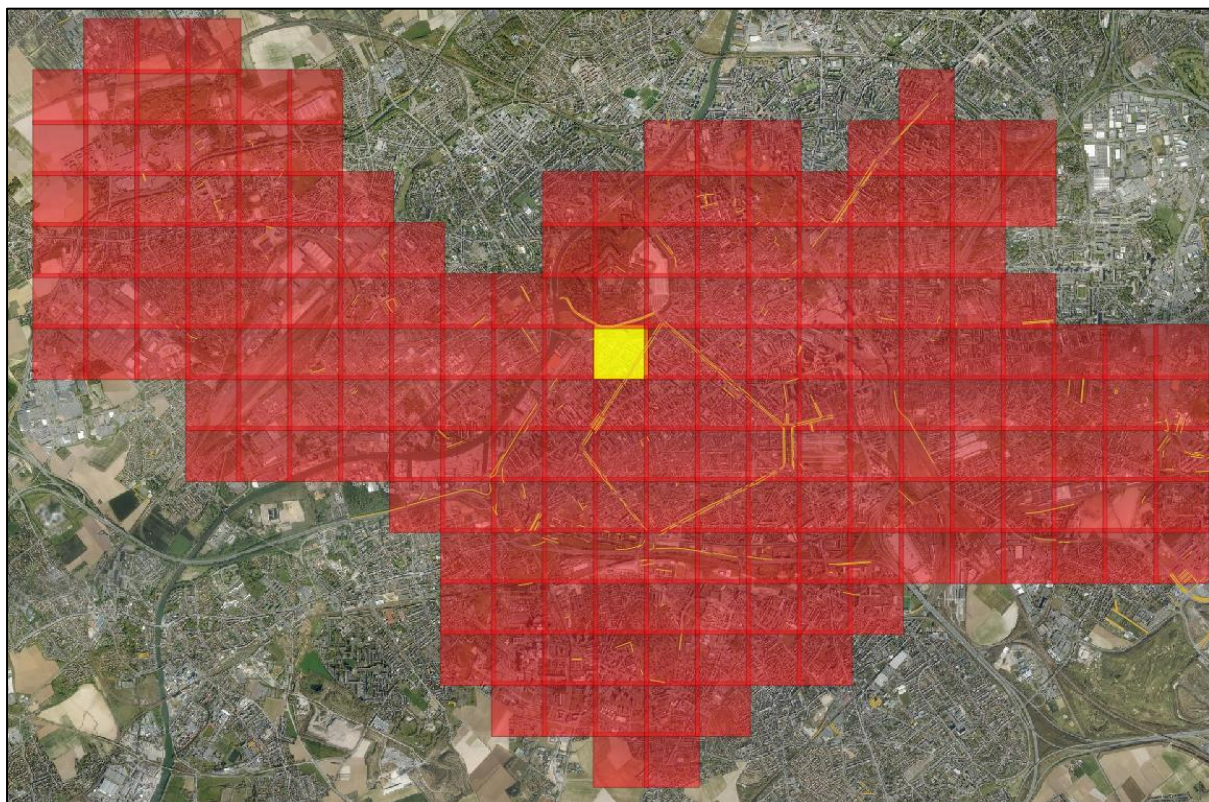


Figure 2 - Tuiles couvrant Lille, Lomme et Hellemmes. En jaune, tuile d'essai.

Pour le test, une seule tuile a été utilisée. Elle a été choisie en fonction de son caractère représentatif et de son utilité : présence de boulevards où un inventaire des arbres existe afin de comparer les résultats prédits et réels, présence de zones boisées et de secteurs privés arborés.

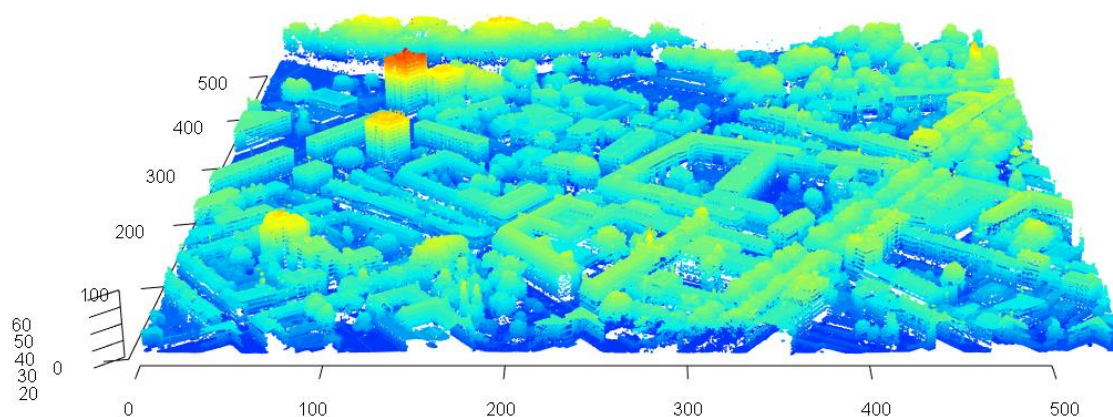


Figure 3 - Tuile de test

Voici les caractéristiques de cette tuile :

Class	LAS (v1.4 format 8)
Memory	2.8 Gb
Extent	702980, 703520, 7059480, 7060020 (xmin, xmax, ymin, ymax)
coord. ref.	RGF93 / Lambert-93

area	285140 m ²
points	36.11 million points
density	126.64 points/m ²

- 3- Chargement de la tuile en éliminant les données aberrantes issues d'erreurs de captures (présence d'oiseaux...)
- 4- Classification du terrain à l'aide la méthode CSF (2) : les points au niveau du sol se voient attribuer une classe afin de les distinguer du reste.

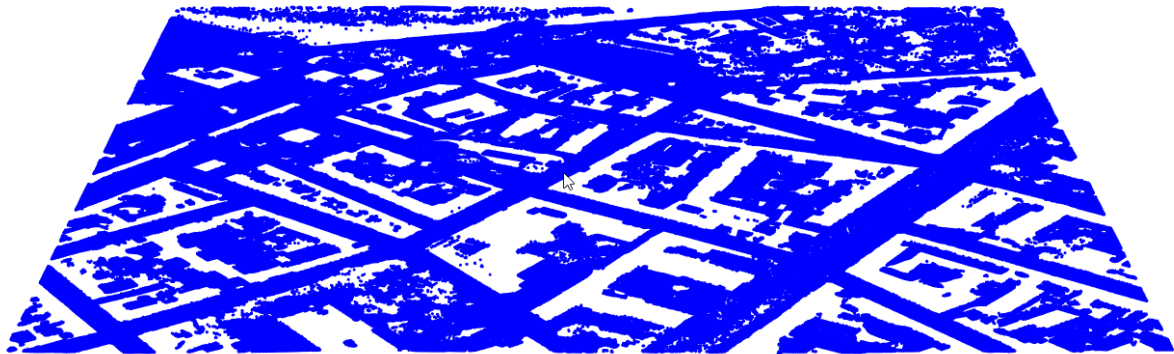


Figure 4 - Nuage de points au niveau du sol

- 5- Normalisation du nuage avec l'algorithme knnidw basé sur le nuage de points : « aplanit » le nuage de points de façon à rendre comparables les données entre elles. Révision décembre : cette étape n'était pas à sa place. Dans la version définitive, celle-ci se situe désormais avant la création du modèle de canopée.
- 6- Suppression des données de faible altitude et du sol : le but est d'éliminer les points inutiles ainsi que les véhicules pour gagner en vitesse de traitement et en fiabilité.
- 7- Chargement de la couche des bâtiments avec les informations de hauteur et suppression de ces informations du nuage de points :

Afin de ne travailler que sur les arbres, il est nécessaire d'exclure les bâtiments du traitement. Pour ce faire, afin d'éviter un autre travail de classification du nuage de points spécifique au bâti et d'avoir les données les plus récentes, j'ai utilisé la base de données topographique de l'IGN. Il est sans doute possible d'utiliser d'autres données.

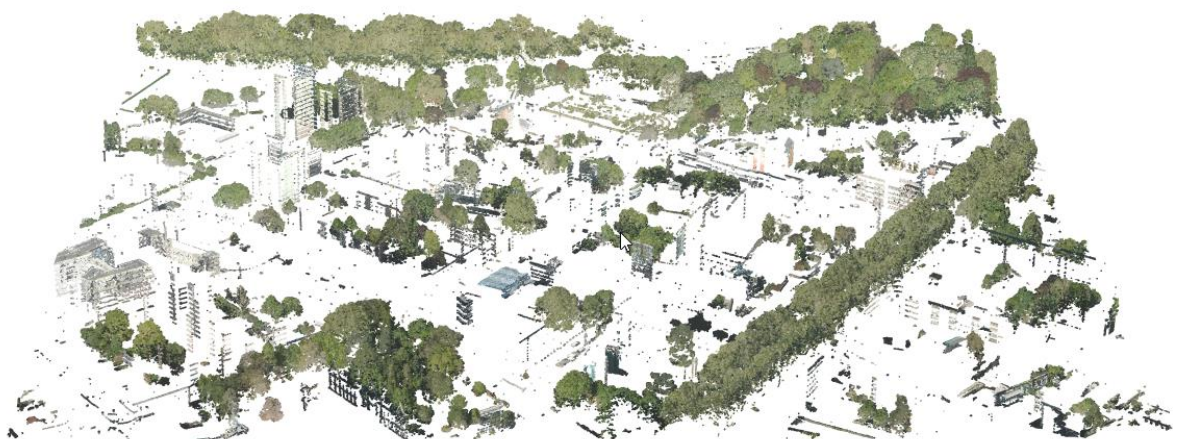


Figure 5 - Suppression des bâtiments avec un tampon d'1 m. Restent apparents quelques bâtiments non présents dans la BD Topo et quelques éléments techniques qui seront filtrés par la suite.

Les bâtiments ont été exclus du nuage de point en fonction de leur hauteur, afin de garder le maximum de points qui pourraient faire partie de la canopée. Cependant, la position des bâtiments souffre d'imprécision. De plus, la hauteur seule ne permet de délimiter que des objets cubiques sans tenir compte des pentes des toitures. En outre, ne sont pas pris en compte les balcons ni les éléments techniques sur les toitures (cheminées et autres).

De ce fait, j'ai ajouté 2 m. à l'emprise latérale des bâtiments et n'ai pris en compte que la hauteur maximale disponible.

Révision décembre : l'utilisation du MNTB mis à disposition par la MEL offre un meilleur filtrage.

- 8- Filtrage complémentaire : à l'aide de fonctions intégrées, ont été supprimés tous les éléments dont la topologie est distincte des arbres, notamment tous les points coplanaires.
- 9- Création du modèle de hauteur de canopée (CHM) à l'aide l'algorithme pit-free (3) et filtrage : sert à délimiter les couronnes des arbres et à leur localisation.

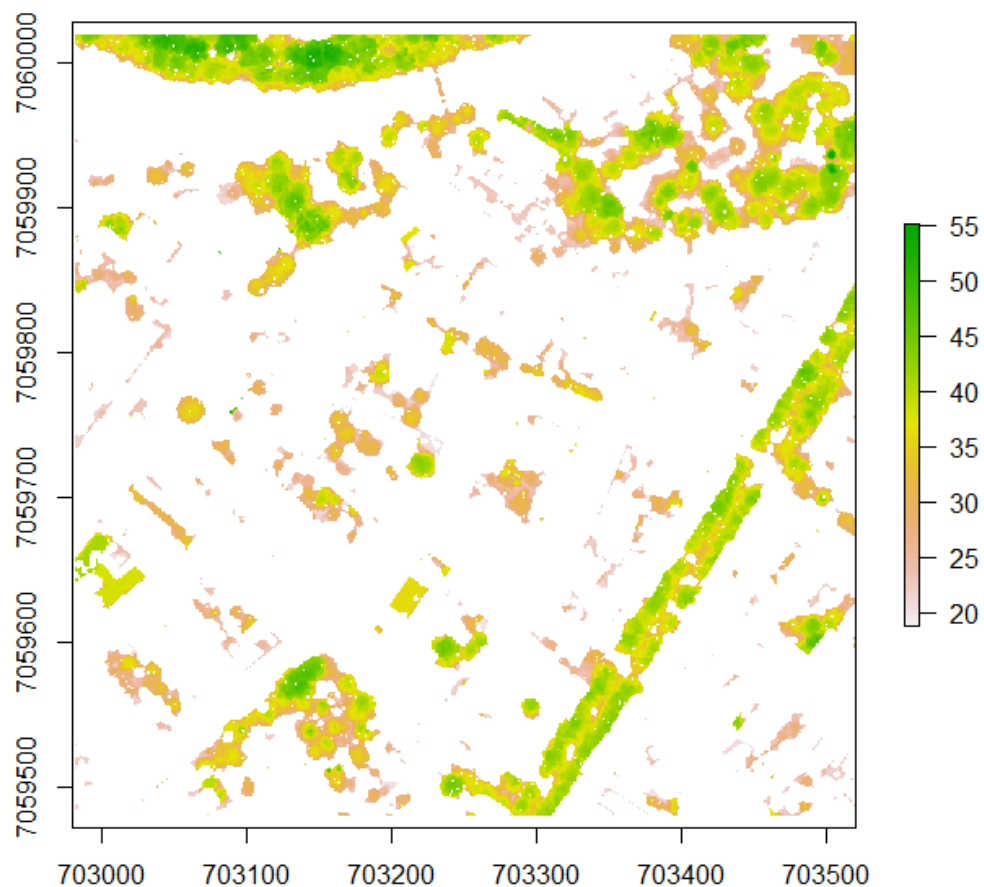


Figure 6 - Le modèle de hauteur de canopée en 2D (fausses couleurs) avec un filtrage léger

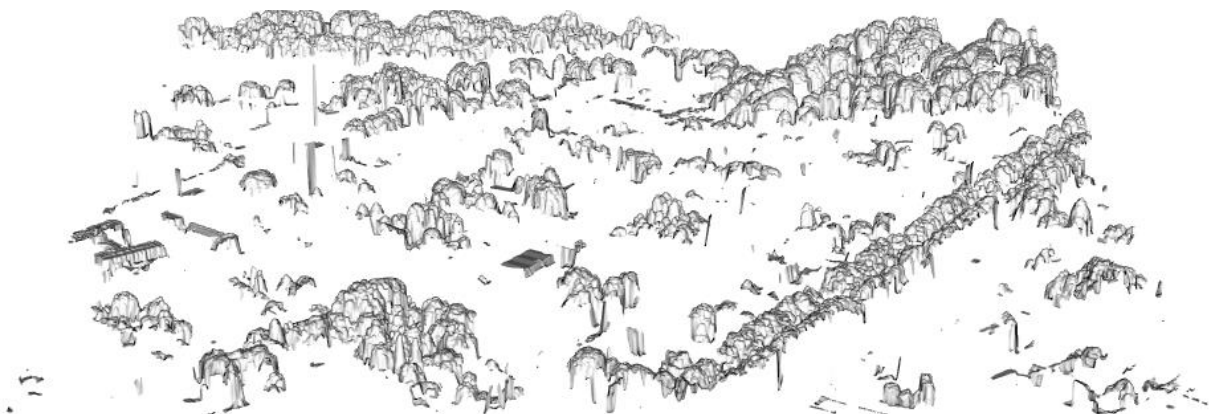


Figure 7 - Le même modèle de hauteur de canopée en 3D

10- Détection individuelle des arbres à l'aide de l'algorithme « Local Maximum Filter » (lmf) sur le nuage de points : une taille de fenêtre (de recherche) variable a été fixée à l'aide d'une fonction afin d'optimiser la vitesse et la surface de recherche. La hauteur minimale pour la recherche a été fixée à 3 m. et l'identification unique de chaque arbre sur « bitmerge ».

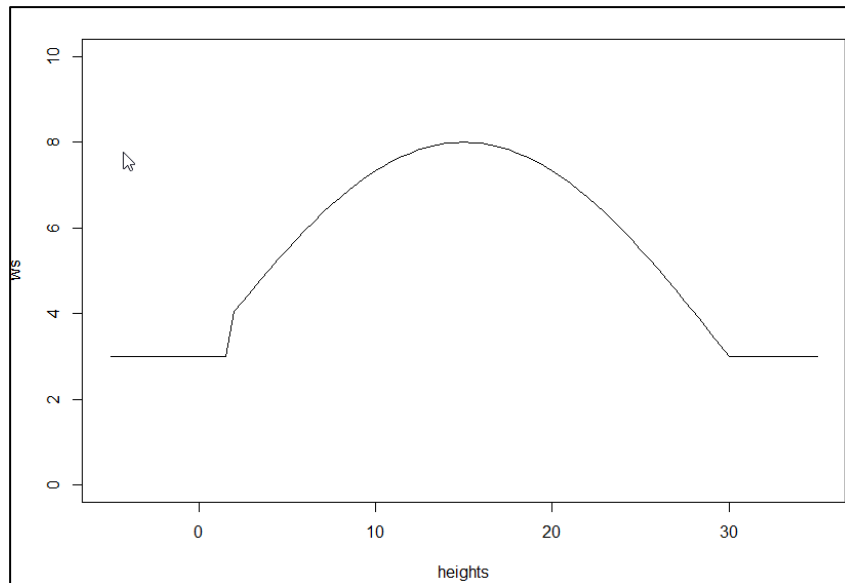


Figure 8 - La fonction définissant la taille de la fenêtre de recherche

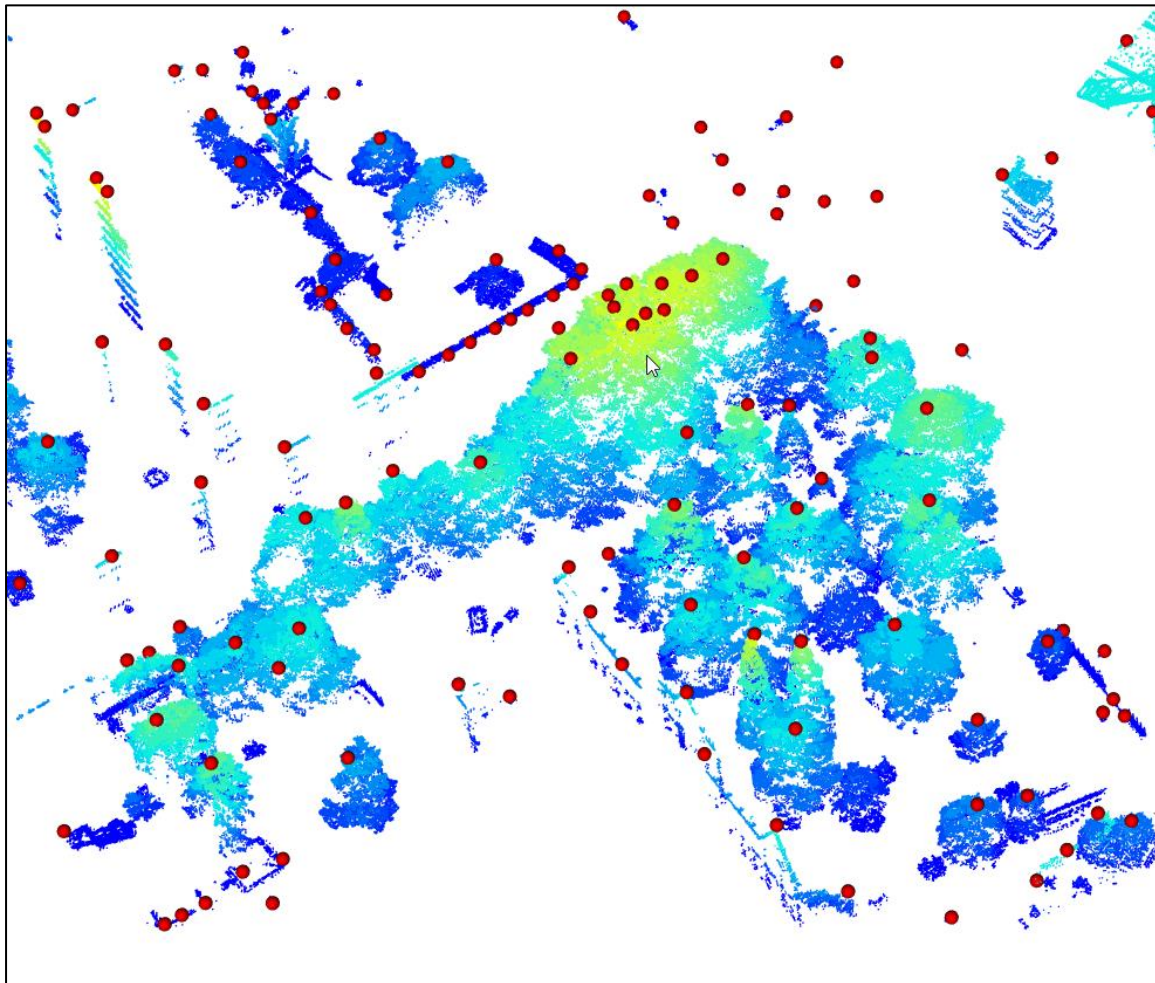


Figure 9 - Illustration du positionnement des crêtes et de plusieurs erreurs de détection dues à de nombreux artefacts

On peut constater sur la figure 9 que de nombreux points « parasitent » le traitement. Ceux-ci sont majoritairement issus de façades et autres éléments de construction non éliminés du nuage de point aux étapes précédentes. A ce stade, on pourrait revenir et améliorer ces points. Cependant ces artefacts pourront être éliminés dans des traitements ultérieurs.

Révision décembre : l'utilisation d'un filtrage basé sur un indice (NDVI...) permet un filtrage efficace des données aberrantes.

- 11- Segmentation du nuage de points : afin de calculer les caractéristiques de chaque arbre, on va créer une entité géographique polygonale sur la base de la couronne arborée qui pourra être exportée dans un fichier facile à lire dans les applications cartographiques.

À cette fin, plusieurs algorithmes sont disponibles, dont watershed, dalponte (4) et li2012 (5). Seul li2012 se base sur le nuage de point, les autres vont utiliser le CHM et la position calculée précédemment. Trois exports ont été générés avec des paramètres donnant de bons résultats (watershed-4, dalponte 150, li2012-3-5).

- 12- Nettoyage des points non attribués à des arbres et classification en fonction de l'identifiant unique.



Figure 10 - Chaque arbre identifié a sa propre couleur. On repère facilement les erreurs de segmentation et de classement. Utilisation de l'algorithme dalponte (4).

- 13- Calculs métriques : chaque entité peut se voir attribuer des données calculées liées au nuage de points qui lui sont propres, par exemple, la hauteur moyenne, la couleur moyenne, l'intensité, l'aplatissement. Ces données peuvent être utilisées par la suite dans un SIG ordinaire afin d'effectuer des filtrages et des calculs complémentaires.
- 14- Export dans un fichier 2D, concavité = 2, avec des métriques variées dont la couleur : nettoyage final et exploitation dans un SIG.

15- Validation des mesures : malgré les nombreuses imprécisions qui nécessiteraient des corrections, il est possible de tester une méthode de validation des mesures. Il suffit de comparer les données réelles mesurées sur le terrain (points verts sur la figure 11) et les données calculées par la chaîne de traitement (polygones) sur un secteur bien précis.

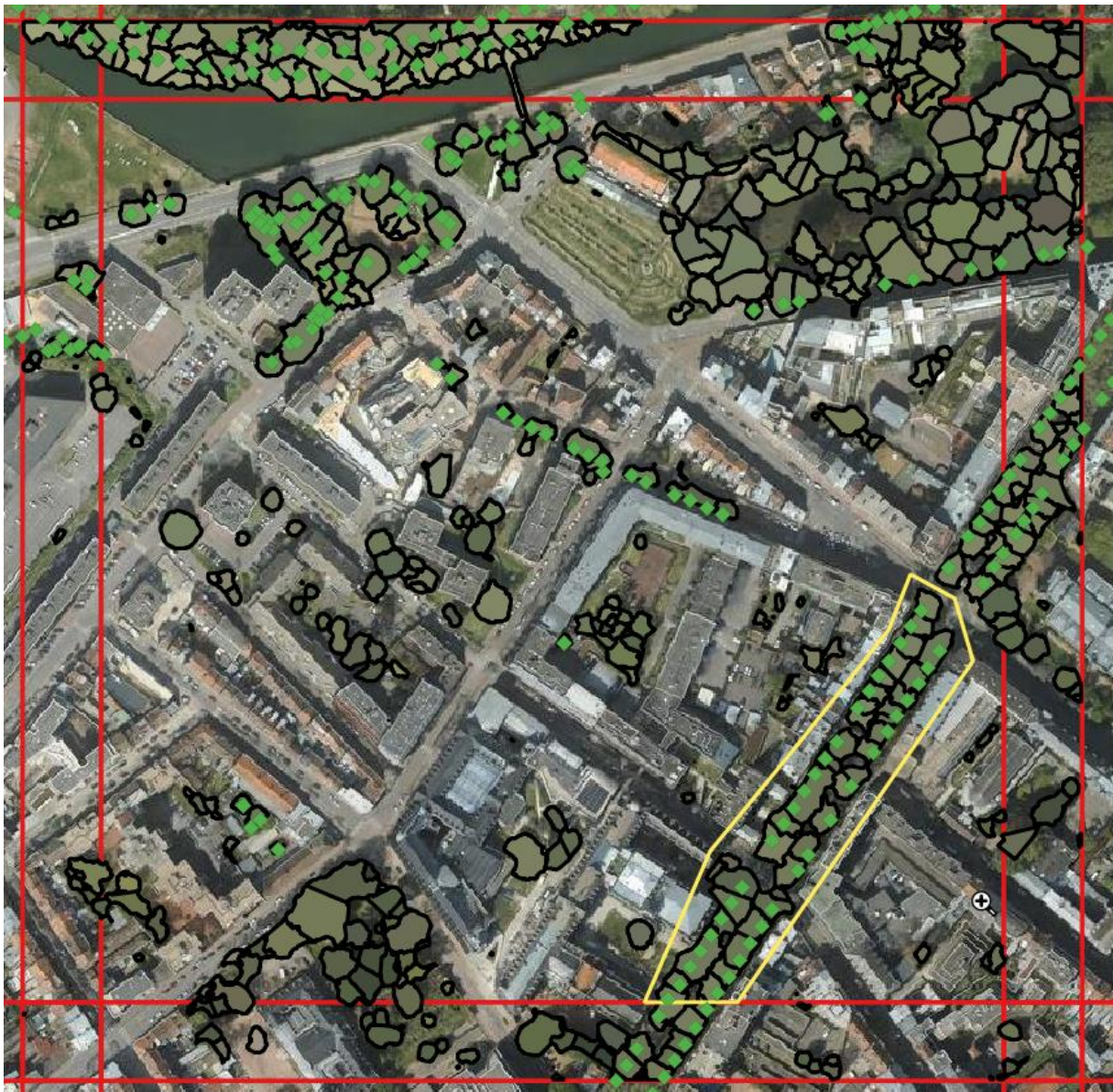


Figure 11 - Import des données dans QGIS sur la photographie aérienne. Algorithme dalponte

Résultat

Ici a été dessiné en jaune, sur le boulevard Vauban, un secteur d'environ 1 ha, sur les 25 ha que compte une tuile. Bien que simple et visiblement très imprécis au niveau de la délimitation des couronnes,

somme_mesuree	39
somme_calculée_watershed	41
somme_calculée_dalponte	67
somme_calculée_li2012	236

Figure 12 - Comparaison des résultats

l'algorithme watershed donne les meilleurs résultats en termes de dénombrement, 41 arbres détectés sur les 39 existants.

Discussion

On peut critiquer le rendu final qui est loin d'être parfait et mériterait des ajustements, à chaque étape du traitement. Rappelons qu'il s'agit d'une expérimentation et qu'il était prévu dès le départ de quantifier la marge d'erreur. Il est grandement possible d'améliorer le résultat, notamment en améliorant l'élimination du bâti et la classification d'autres éléments perturbateurs, tels les véhicules. En outre, la chaîne de traitement pourrait être revue avec les conseils avisés de l'auteur du package lidR et l'appui de sigistes confirmés dans le traitement de données LiDAR.

Révision décembre : quelques conseils ont été donnés par l'auteur, la chaîne de traitement fournie en tient compte.

La durée du traitement ne dépasse pas 20 minutes pour une tuile, soit une durée totale de traitement d'un peu moins de trois jours pour les 195 tuiles couvrant Lille, Lomme et Hellemmes.

Révision décembre : quelques arrêts peuvent subvenir à différents endroits du traitement qui peuvent nécessiter un traitement manuel. La chaîne fournie n'a rencontré qu'un seul arrêt sur une tuile dont quelques données LiDAR étaient aberrantes.

Actuellement, la méthode reste incomplète. Elle ne permet pas d'obtenir une information précise dans certains cas : arbres trop petits, couverts ou mêlés de façon trop dense à d'autres arbres. Cependant, elle semble offrir des données d'une qualité supérieure aux autres méthodes.

En outre, les mesures dérivées n'ont pas été poussées assez loin pour obtenir des informations utiles sur les essences ou sur les capacités d'absorption des polluants, par exemple. Ceci pourrait faire l'objet d'études complémentaires. De même une méthode de comptage manuelle très précise basée sur des coupes et la mise en exergue des troncs pourrait faire l'objet d'une étude partagée avec les gestionnaires.

Conclusion

Bien qu'incomplète, cette expérience montre qu'il est possible d'obtenir des informations sur le nombre d'arbres et d'autres informations utiles, comme la surface de canopée ou la hauteur moyenne de chaque arbre du patrimoine lillois avec l'outil lidR.

Cependant, le travail n'est pas terminé, il reste à affiner la méthode et à corriger, étape par étape, les différents problèmes de mesure et de calibration.

Il aura cependant permis de confirmer l'importance du travail de terrain et la nécessité d'avoir des données de référence fiables. Il ressort également qu'un identifiant unique permettant de suivre la vie d'un arbre est important, tant en terme de gestion que de recollement des données.

Références

- 1- ROUSSEL, JEAN-ROMAIN & AUTY, DAVID & COOPS, NICHOLAS & TOMPALSKI, PIOTR & GOODBODY, TRISTAN & SÁNCHEZ MEADOR, ANDREW & BOURDON, JEAN-FRANÇOIS & DE BOISSIEU, FLORIAN & ACHIM, ALEXIS. (2020). LIDR: AN R PACKAGE FOR ANALYSIS OF AIRBORNE LASER SCANNING (ALS) DATA. REMOTE SENSING OF ENVIRONMENT. 251. 112061. 10.1016/j.rse.2020.112061.

- 2- ZHANG, W.; QI, J.; WAN, P.; WANG, H.; XIE, D.; WANG, X.; YAN, G. AN EASY-TO-USE AIRBORNE LIDAR DATA FILTERING METHOD BASED ON CLOTH SIMULATION. *REMOTE SENS.* 2016, 8, 501. 10.3390/rs8060501
- 3- KHOSRAVIPOUR, ANAHITA & SKIDMORE, ANDREW & ISENBURG, MARTIN & HUSSIN, YOUSIF. (2014). GENERATING PIT-FREE CANOPY HEIGHT MODELS FROM AIRBORNE LIDAR. *PHOTOGRAMMETRIC ENGINEERING & REMOTE SENSING.* 80. 863-872. 10.14358/PERS.80.9.863.
- 4- DALPONTE, M. AND COOMES, D.A. (2016), TREE-CENTRIC MAPPING OF FOREST CARBON DENSITY FROM AIRBORNE LASER SCANNING AND HYPERSPECTRAL DATA. *METHODS ECOL EVOL*, 7: 1236-1245. 10.1111/2041-210X.12575
- 5- LI, W., GUO, Q., JAKUBOWSKI, M.K., KELLY, M., 2012. A NEW METHOD FOR SEGMENTING INDIVIDUAL TREES FROM THE LIDAR POINT CLOUD. *PHOTOGRAMM. ENG. REMOTE SENS.* 78 (1), 75–84