



Apport des données LiDAR brutes de l'IGN pour l'amélioration des inventaires de végétation de SNCF Réseau

■ Samuel LUON - Fatoumata BAH - Anne-Sophie ONODY

Pour répondre aux enjeux de maintenance, SNCF Réseau a besoin de connaître le patrimoine végétal des emprises ferroviaires. Pour cela, une chaîne de traitement automatisée permettant d'identifier la végétation et de la classer en trois strates (arborée, arbustive, herbacée) grâce à un algorithme de classification supervisée Random Forest a été développée. Les résultats de cette classification montrent certaines confusions entre les classes arborées et arbustives. Pour améliorer les résultats, la piste d'utilisation de données de hauteur de la végétation issue du traitement des données LiDAR de l'IGN a été testée. Cette amélioration passe par la création d'un Modèle Numérique de Hauteur (MNH) issu des données LiDAR. Pour rendre cette démarche opérationnelle et répondre aux besoins d'industrialisation, la méthodologie a été développée en utilisant le langage de programmation R et tirant parti de la librairie lidR.

MOTS-CLÉS

LiDAR HD, MNH, classification, strates, végétation, nuage de points

présente diverses faiblesses, d'une part l'impossibilité d'exploiter des images avec une couverture nuageuse, d'autre part, le maintien d'une confusion entre les trois strates et enfin, l'imprécision du paramètre de hauteur (Viguiier 2015). Le LiDAR se démarque par sa capacité à fournir des données tridimensionnelles de haute résolution, de capturer la structure verticale de la végétation offrant une analyse approfondie de la répartition spatiale des différentes strates végétales.

Méthodologie

■ Zone d'étude

Pour cette étude, deux tronçons de deux lignes ferrées distinctes situées dans la région Auvergne-Rhône-Alpes, où les données LiDAR HD brutes sont disponibles et ont été analysées.

■ Présentation du programme LiDAR HD

Le programme LiDAR HD est un programme coordonné par l'IGN (Institut national de l'information géographique et forestière). LiDAR HD a pour objectif d'acquérir des données LiDAR Haute Densité (HD) sur l'ensemble du territoire français pour répondre aux besoins en analyse spatiale dans divers domaines de l'action publique, foresterie, prévention des risques...

Les données sont mises à disposition de tous, gratuites et différents produits sont proposés : les données LiDAR brutes, les données LiDAR classées, des MNT, MNS et MNH au pas de 50 cm, 1m et 5 m. Début 2024, les produits énoncés ne sont pas tous disponibles. Les données LiDAR brutes ou les données classées sont uniquement disponibles sur certaines parties du territoire et les différents modèles numériques sont à ce jour indisponibles.

Introduction

SNCF Réseau est le gestionnaire du réseau ferré national et a donc la charge du développement des infrastructures et de leur maintenance, au total, ce sont 95 000 hectares d'emprises ferroviaires à entretenir. En conséquence, il est important de connaître la situation des zones de voies et de préparer les moyens adéquats pour les opérations de maintenance. Parmi ces nombreuses interventions, on retrouve les opérations de maîtrise de la végétation. En effet, la végétation constitue une source potentielle de dysfonctionnements. Les chutes d'arbres en particulier restent le risque majeur, car elles peuvent stopper tout trafic.

Afin de réduire ces potentiels risques, SNCF Réseau a décidé d'ajuster au mieux la programmation des travaux en fonction de l'état de la végétation. Dans cette optique, l'entreprise a donc mis en place, en 2018, un inventaire du patrimoine végétal le long des lignes appartenant à SNCF Réseau. Cet inventaire annuel est classé en quatre catégories (arborée, arbustive, herba-

cée et absence de végétation) et est obtenu par traitement d'image satellite Pléiades. Cependant, la distinction entre les strates arborée et arbustive demeure complexe, principalement pendant la saison végétative, lorsque la présence de feuilles rend difficile leur identification par images satellite.

La hauteur est un critère discriminant permettant de les distinguer. La technologie LiDAR apparaît alors comme une bonne solution, et cela, surtout depuis le programme LiDAR HD de l'IGN qui permet d'obtenir de façon libre et gratuite des données LiDAR sur la France entière. Les données LiDAR permettent de générer différents modèles numériques, dont le modèle numérique de hauteur (MNH) qui représente la différence d'altitude entre le Modèle Numérique de Surface (MNS) et le Modèle Numérique de Terrain (MNT). D'autres technologies de télémétrie permettent de mesurer des hauteurs comme le RADAR, la photogrammétrie ou la tri-stéréoscopie. Toutefois, aucune ne présente un niveau de précision de l'ordre du centimètre, similaire au LiDAR. Par exemple, la tri-stéréoscopie



► **La donnée LIDAR HD et le téléchargement**

Au moment de l'étude, la seule donnée LiDAR HD disponible était la donnée brute. En juin 2023 sont sorties les premières données classifiées.

Le LiDAR utilisé par l'IGN est un système LiDAR-ALS de type multi-échos utilisant un laser infrarouge ("LiDAR HD : vers une nouvelle cartographie 3D du territoire - Institut - IGN", s. d.). Il est donc capable d'enregistrer un certain nombre d'échos pour chaque tir laser et de fournir des informations sur la structure verticale de la végétation, dans la limite de ses caractéristiques techniques et du type de végétation. L'IGN réalise un contrôle afin d'assurer une précision minimum de : 50 cm d'EMQ (erreur moyenne quadratique) en planimétrie et 10 cm d'EMQ en altimétrie, il s'assure également que les nuages de points présentent une densité d'au moins 10 impulsions au m². L'information de la classe ne concerne que les données LiDAR HD classées, les données brutes, quant à elles, ne peuvent contenir qu'une seule classe (0 ou 1). La classe sol est l'une des classes les plus importantes, elle est indispensable à la création de MNT, MNS et MNH. Elle déterminera également la précision des différents modèles numériques. Pour les données brutes, la classe sol étant inexistante, il sera donc primordial de classer le sol afin de pouvoir générer un MNH.

Pour le téléchargement, un script R, un langage de programmation spécialisé dans l'analyse statistique et la manipulation de données, a été créé afin de télécharger automatiquement les données nécessaires. Les données LiDAR sont des données lourdes qui occupent beaucoup d'espace. Dans une éventualité de télécharger les données à l'échelle nationale pour le réseau ferré, optimiser l'espace de stockage est essentiel et cela passe tout d'abord par la suppression des données inutiles jusqu'à la génération et l'utilisation

Format	.LAS	.LAZ
Taille du fichier	1 012 052 ko	189 862 ko

Tableau 1. Comparaison de la taille d'un fichier au format LAS et LAZ.

uniquement de fichiers .laz (tableau 1) qui est un format compressé.

► **Présentation de la méthodologie de calcul du MNH à partir du nuage de points LiDAR**

Le créateur de la librairie lidR (Roussel et al. 2020) indique qu'il y a des traitements récurrents pour traiter des données LiDAR ALS. Ces traitements comprennent, dans l'ordre, la classification du sol, la génération d'un modèle numérique de terrain et d'un modèle numérique de surface.

Classification du sol via lidR

Cette étape est une étape critique, car elle conditionne le résultat et la précision des MNT, MNS et MNH. La majorité des algorithmes utilisés pour la classification des retours au sol utilisent des filtres morphologiques et spatiaux. La librairie lidR propose trois algorithmes pour classer le sol : PMF (*Progressive Morphological Filter*), CSF (*Cloth Simulation Filter*), MCC (*Multiscale Curvature Classification*). Pour se rendre compte du résultat de la classification, des coupes transversales ont été créées sur R (figure 1).

La comparaison des résultats montre que, malgré les paramètres par défaut, l'algorithme PMF ne parvient pas à bien discriminer le sol des éléments non-sol, tandis que CSF et MCC réalisent une meilleure classification. Bien que MCC puisse fournir une classification satisfaisante, l'optimisation des paramètres par défaut de CSF s'avère utile pour traiter efficacement des volumes de données représentant différents types de terrains. Outre la précision de la classification, le temps de traitement est également crucial (tableau 2). Ainsi, CSF est privilégié pour la classification du sol en raison de sa performance temporelle et de sa précision.

Génération du MNH

LidR offre la possibilité de créer des MNT, MNS et MNH. Généralement le MNH est issu de la combinaison du MNT et du MNS. Cependant, lidR va plus loin en proposant une fonction de normalisation de la surface. L'introduction de cette fonctionnalité réduit la nécessité de générer un MNS distinct. En optant pour la normalisation de la surface, on élimine l'influence

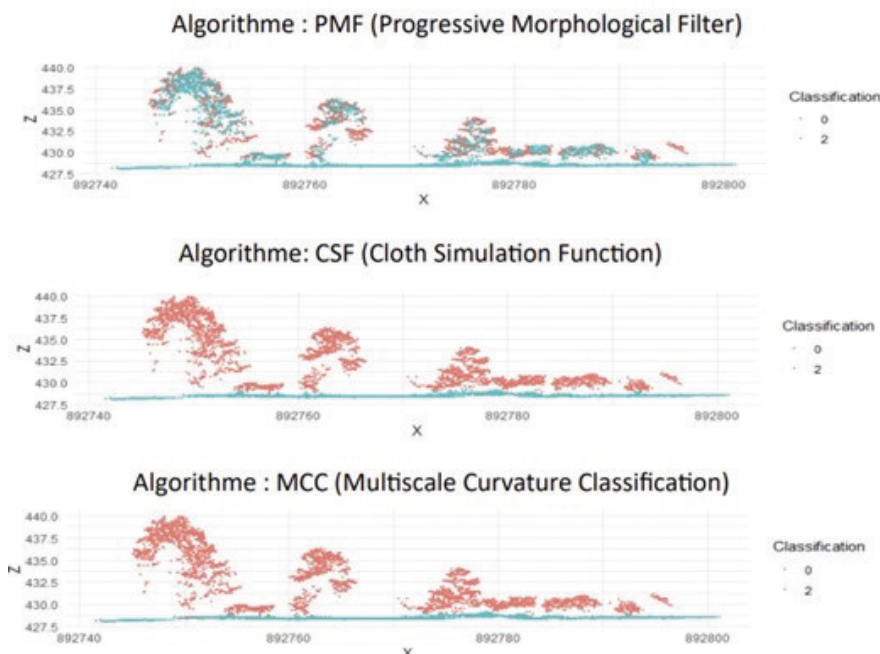


Figure 1. Classification du sol (en bleu) et non-sol (en rouge) sur un nuage de points brut, en fonction des algorithmes PMF, CSF et MCC.

Algorithme de classification du sol	PMF	CSF	MCC
Temps de traitement (s)	154	58.34	525

Tableau 2. Temps de traitement pour un semis selon les algorithmes (Configuration pc : Windows, Intel(R) Xeon(R) W-2123 CPU @ 3.60GHz, 3600 MHz, 4 cœurs, 8 processeurs, RAM : 64Go).



topographique des mesures au-dessus du sol. Techniquement, la normalisation sur lidR consiste à soustraire le MNT d'un nuage de points pour créer une donnée avec un sol à 0 m. Le MNT peut être fourni, mais aussi calculé à la volée, les algorithmes proposés lors de la normalisation sont les mêmes que lors de la création du MNT (TIN, IDW, Krigeage) (Roussel et al. 2020).

Il existe trois algorithmes pour créer un MNH avec les données LiDAR normalisées : point à raster, TIN et Pit-free. L'algorithme choisi pour la création du MNH est le point à raster (p2r). Sa simplicité et sa rapidité d'exécution en font un outil capable de traiter de gros volumes de données.

Afin de généraliser l'application de ce principe à l'ensemble des fichiers LAZ, l'utilisation de la fonction LASCatalog se révèle essentielle. Cette fonctionnalité offre la flexibilité de gérer et d'appliquer divers traitements à une collection de fichiers LAS ou LAZ, tout en évitant de surcharger la mémoire de la machine. Bien que toutes les fonctionnalités ne soient pas directement compatibles avec LASCatalog, il convient de souligner que la classification du sol, la normalisation et la dérivation en raster MNH peuvent être mises en œuvre de cette manière.

Résultats

Validation du MNH

Afin de rigoureusement valider le MNH, une étape de vérification sur le terrain a été réalisée le long du tronçon de la ligne Lyon-Marseille en mesurant une cinquantaine de hauteurs de nature différentes : végétation et bâti. Les statistiques montrent que pour la classe "bâtie", l'EMO est de 1,01

m, cela signifie que les prédictions du MNH présentent en moyenne une erreur d'environ 1,01 m par rapport aux hauteurs mesurées sur le terrain. Il est cependant difficile d'interpréter ce résultat sans le replacer dans le contexte. Par exemple, si les hauteurs des bâtiments varient de quelques mètres, alors une erreur de 1,01 m pourrait être considérée comme importante. En revanche, si les bâtiments sont beaucoup plus grands, cette erreur pourrait être plus acceptable. Ainsi, une normalisation de l'EMO est souvent recommandée. La méthode la plus fréquente est de diviser l'EMO par l'étendue des valeurs observées. En normalisant l'EMO, celui-ci est de 0,1 m. Cela suggère que l'erreur moyenne entre les mesures de hauteur diffère de 0,1 m par rapport aux valeurs issues du terrain. Cela peut alors être considéré comme une bonne précision du MNH calculé à partir des données LiDAR HD de l'IGN pour les éléments bâtis.

Pour la classe végétation, l'EMO est de 6,07 m. Cette valeur élevée trouve son origine dans la sensibilité de l'EMO aux valeurs aberrantes. Cependant, en normalisant par l'étendue l'EMO est ramené à 0,22 m. Cette normalisation met en lumière des erreurs moyennes modestes dans les prédictions du MNH par rapport aux mesures de terrain.

En complément de ces analyses, un coefficient de détermination (R^2) a été calculé à partir des régressions linéaires. Pour la classe végétation, le R^2 s'établit à 0,54. Cette valeur indique que le modèle de régression linéaire explique environ 54 % de la variabilité des hauteurs observées pour les objets de la classe végétation, donc la relation entre les hauteurs prédites et observées est modeste. Cette différence significative par rapport à la classe bâtie pourrait

s'expliquer par la nature de l'objet, la hauteur de la végétation étant variable en fonction de facteurs extérieurs, au contraire du bâti qui a une hauteur fixe au cours du temps. Cela suggère aussi que le décalage temporel entre l'acquisition des données LiDAR utilisées pour construire le modèle numérique de hauteur et les mesures de terrain a un impact sur la qualité du modèle de régression. Si la végétation a connu une croissance significative au cours de cette période, les données LiDAR plus anciennes ne tiendront pas compte de cette variation, ce qui pourrait expliquer la réduction de l'ajustement du modèle aux données actuelles.

En tenant compte de la validation sur le terrain et des différentes statistiques analysées, la méthodologie de création et le MNH ont été validés de manière satisfaisante. L'étape suivante consiste à inclure le MNH dans le processus de classification.

Apport du MNH dans les classifications

SNCF Réseau a mis en place au cours des dernières années une chaîne de traitement automatisée pour la classification d'images satellite Pléiades permettant de distinguer la végétation et de la classer en trois strates distinctes (Onody et al. 2023). L'objectif de cette étude est d'ajouter le MNH calculé à partir des données LiDAR de l'IGN comme variable d'entrée de l'algorithme de classification. Pour évaluer l'impact du MNH obtenu à partir des données LiDAR, une première classification a été effectuée sans utiliser les données du MNH, tandis que la seconde classification a été réalisée en intégrant le MNH. Une photographie acquise par un capteur embarqué

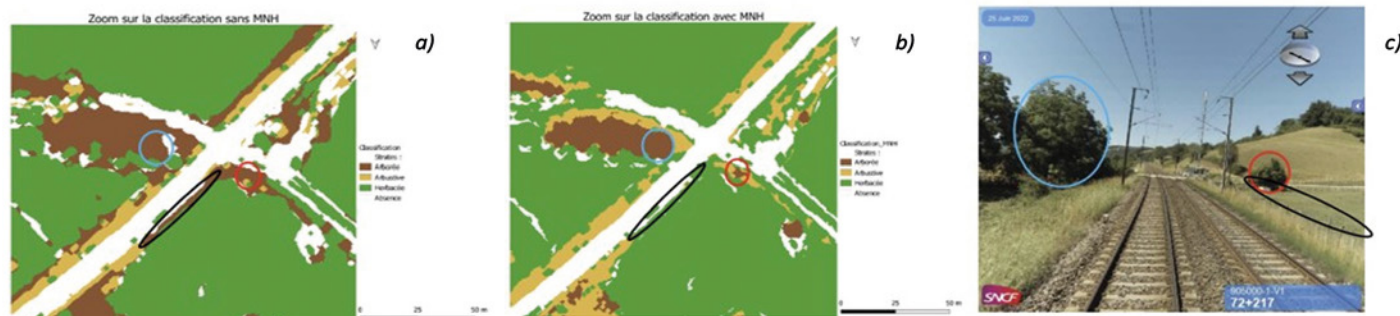


Figure 2. Comparaison classification sans MNH (a) / avec MNH (b) avec Imajnet (c).

Classification sans MNH	Arborée	Arbustive	Herbacée	Absence
Arborée	8	0	1	1
Arbustive	4	4	0	2
Herbacée	3	0	14	3
Absence	0	0	3	7
Précision globale (%)	66			

Classification avec MNH	Arborée	Arbustive	Herbacée	Absence
Arborée	9	0	1	0
Arbustive	0	8	1	1
Herbacée	0	0	17	3
Absence	0	0	3	7
Précision globale (%)	82			

Tableau 3. Matrices de confusion avant et après apport du MNH.

dans la cabine du train faisant office de vérité terrain permet ensuite de faire la comparaison (figure 2).

La figure met en évidence l'impact du MNH sur la distinction des différentes strates de végétation, tout en réduisant les erreurs de classification. Une illustration claire de cet effet est visible dans la zone en noir sur la figure 2a, où la classification initiale suggère la présence d'une strate arborée, tandis que la classification avec l'ajout du MNH indique une strate herbacée (figure 2b). Cette discordance est corroborée par la figure 2c, qui présente une photographie du terrain et confirme la présence de végétation herbacée dans cette zone. Cette comparaison visuelle met en évidence la capacité du MNH à corriger des erreurs de classification initiales. De plus, les zones marquées en rouge et en bleu sur ces figures démontrent une délimitation beaucoup plus précise de la couronne des arbres et donc de la strate arborée.

Outre la manière visuelle, des matrices de confusions ont été créées pour quantifier les précisions des deux classifications. Les deux matrices montrent que la précision globale a été améliorée avec l'apport du MNH passant d'une précision globale de 0,66 à 0,82, soit un

gain de 0,16 (tableau 3). Cette amélioration de 0,16 indique que la classification avec le MNH a une capacité accrue à attribuer correctement les classes de végétation par rapport à la classification initiale. Une autre classification a été réalisée dans une zone différente de l'espace d'étude et elle montre une légère amélioration de la précision après l'ajout du paramètre de hauteur. Contrairement à la première comparaison, le gain est léger. La classification sans l'intégration du MNH présente une précision de 0,71, tandis qu'avec l'ajout du MNH, la précision passe à 0,73. Cette légère amélioration est peut-être due aux caractéristiques paysagères de la zone d'étude. Cette deuxième zone étudiée est située dans un contexte urbain avec peu de zones boisées et donc moins de strates arborée et arbustive à identifier.

Jusqu'à présent, les MNH utilisés se sont basés sur des données LiDAR acquises en été, ce qui les a mis en concordance saisonnière avec les images satellite Pléiades utilisées pour les classifications. Cependant, il est à noter qu'il existe des acquisitions LiDAR effectuées en hiver, dans les régions situées au nord et à l'est de la France. Cette différence saisonnière dans les

acquisitions LiDAR et les acquisitions Pléiades acquises par SNCF Réseau pour la génération de ses inventaires du patrimoine végétal a un impact significatif sur les classifications (figure 3). On observe sur la figure 3b que le MNH calculé sur les données LiDAR acquises en hiver sous-estime la hauteur de la végétation, notamment dans les zones forestières. Cela entraîne alors des confusions dans la classification *Random Forest*. Ces erreurs de MNH sont certainement liées à l'essence des arbres dont la densité des feuilles, moins importante en hiver, laisse passer plus facilement les faisceaux laser du LiDAR (Air 2023).

Discussion

L'objectif central de cette étude était d'évaluer dans quelle mesure les données LiDAR HD brutes de l'IGN pouvaient améliorer le système de classification de SNCF Réseau. Les résultats obtenus à travers les différentes classifications indiquent de manière générale que l'intégration du MNH, créé à partir des données LiDAR, au processus de classification, conduit à des améliorations dans la précision de la classification. Toutefois, il est

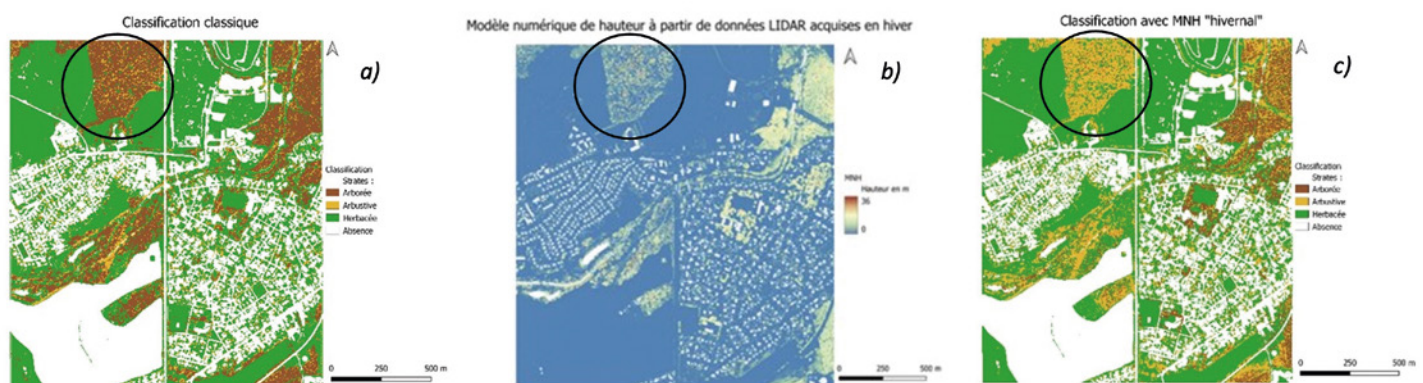


Figure 3. Effet d'un MNH acquis en hiver.



important de noter que l'ampleur de cette amélioration varie en fonction du type de terrain étudié. Par ailleurs, il a été démontré que les données LiDAR HD acquises en hiver ne répondent pas à la problématique, principalement en raison du décalage saisonnier entre les deux sources de données, à savoir les données LiDAR et les images satellite. Cette question de temporalité se pose de manière plus générale dans le contexte où les données LiDAR sont acquises uniquement une seule fois, ainsi chaque espace du territoire représente une situation à un moment T précis. Contrairement à cela, les images satellite utilisées par SNCF Réseau sont acquises chaque année, fournissant ainsi une mise à jour annuelle de l'information, prenant donc en compte les changements du terrain (croissance de la végétation, abattage, etc.). Cette disparité dans la fréquence des acquisitions entre les données LiDAR et les images satellite peut entraîner des incohérences dans la classification, car les caractéristiques du terrain peuvent évoluer au fil du temps. Il est donc essentiel de tenir compte de cette différence de temporalité lors de l'intégration des données LiDAR dans le processus de classification afin de garantir des résultats fiables. À long terme, se pose aussi la question de l'obsolescence des données LiDAR HD et de leur pertinence pour les besoins des classifications. Les besoins en informations actualisées et précises pour la classification, en particulier dans les zones sujettes à des changements fréquents, nécessitent une réflexion sur la manière de maintenir la pertinence des données LiDAR issues du programme LiDAR HD si celles-ci ne sont pas renouvelées après 2026.

Conclusion

En conclusion, la méthodologie mise en place a permis de constater que l'ajout du MNH, issu des données LiDAR HD, dans le processus de classification a permis d'améliorer la qualité des inventaires. Ce résultat est particulièrement notable dans les contextes où les acquisitions LiDAR sont réalisées avec une concordance temporelle avec les

images satellite utilisées pour les classifications. Au contraire, les données LiDAR ne permettent pas d'obtenir l'information souhaitée lorsque celles-ci sont acquises en hiver. Cette amélioration dans les classifications permet à SNCF Réseau d'optimiser et affiner sa connaissance de son patrimoine végétal, ce qui permet une meilleure gestion et préservation du réseau ferré. Cette donnée gratuite constitue pour la SNCF une mine de renseignements et peut être employée pour diverses problématiques. Ce travail exploratoire ne s'arrête donc pas ici et les perspectives sont encore nombreuses, allant de l'identification et l'individualisation des arbres jusqu'au repérage des arbres à risques.

De plus, les traitements sont industrialisables, ce qui permet de reproduire et d'utiliser la méthodologie employée à l'échelle nationale. Cela nécessite toutefois des capacités de stockage importantes. L'IGN, dans le futur, prévoit la diffusion des différents modèles numériques dont le MNH. Cette avancée permettra sans nul doute de gagner en temps de traitements et en capacité de stockage, mais aussi de renforcer la robustesse des classifications. ●

Références

- Air, Drone On. 2023. "ACTU-TECH - LiDAR : PRINTEMPS vs HIVER". DRONE ON AIR (blog). 15 mai 2023. <https://www.droneonair.com/actu-tech-lidar-printemps-vs-hiver/>.
- "LiDAR HD : vers une nouvelle cartographie 3D du territoire - Institut - IGN". s. d. Consulté le 18 janvier 2024. <https://www.ign.fr/institut/lidar-hd-vers-une-nouvelle-cartographie-3d-du-territoire>.
- Onody, Anne-Sophie, Fatoumata Bah, Marie-Ange Boum, et Grégory Duval. 2023. "Use of satellite imagery to categorize vegetation on the French railway network (SNCF Réseau)". *Transportation Research Procedia, TRA Lisbon 2022 Conference Proceedings Transport Research Arena (TRA Lisbon 2022)*, 14th-17th November 2022, Lisboa, Portugal, 72 (janvier): 1451-58. <https://doi.org/10.1016/j.trpro.2023.11.610>.
- Roussel, Jean-Romain, David Auty, Nicholas C. Coops, Piotr Tompalski, Tristan R. H. Goodbody, Andrew Sánchez Meador, Jean-François Bourdon, Florian de Boissieu, et

Alexis Achim. 2020. "lidR: An R package for analysis of Airborne Laser Scanning (ALS) data". *Remote Sensing of Environment* 251 (décembre): 112061. <https://doi.org/10.1016/j.rse.2020.112061>.

Vigier, Flavien. 2015. "Le contrôle de la végétation dans les emprises ferroviaires : une approche multi-scalaire". *XYZ* N° 144, septembre 2015, pp. 25-40.

Contacts

Luon Samuel (SNCF)
ext.samuel.luon@reseau.sncf.fr
Onody Anne-Sophie (SNCF)
anne-sophie.onody@reseau.sncf.fr
Bah Fatoumata (SNCF)
fatoumata.bah@reseau.sncf.fr

ABSTRACT

Keywords: LiDAR HD, DHM, classification, stratum, vegetation, point cloud

To meet the challenges of maintenance, SNCF Réseau needs to know what vegetation is growing on railway rights of way. To do this, an automated processing chain was developed to identify the vegetation and classify it into three strata (trees, shrubs, and herbaceous plants) using a Random Forest supervised classification algorithm. The results of this classification show some confusion between the tree and shrub classes. To improve the results, the use of vegetation height data derived from IGN LiDAR data processing has been explored. This improvement involves the creation of a Digital Height Model (DHM) based on LiDAR data. To make this approach operational and meet the need for industrialization, the methodology was developed using the R programming language taking advantage of the lidR library.

NDLR : des raisons éditoriales font que l'article annoncé dans le précédent numéro 177, la "campagne de mesures inédites pour l'étude des bienfaits des arbres sur le climat de trois rues à Strasbourg" est absent de ce numéro. Que le lecteur ne s'impatiente pas et attende le prochain numéro XYZ 179 de juin.