



HAL
open science

Fusion de données LiDAR aéroporté et SENTINEL-2 par apprentissage pour la caractérisation de la ressource forestière du PNR des Bauges

Kamel Lahssini

► **To cite this version:**

Kamel Lahssini. Fusion de données LiDAR aéroporté et SENTINEL-2 par apprentissage pour la caractérisation de la ressource forestière du PNR des Bauges. Sciences de l'environnement. 2021. dumas-03575198

HAL Id: dumas-03575198

<https://dumas.ccsd.cnrs.fr/dumas-03575198v1>

Submitted on 15 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0
International License

AgroParisTech - Montpellier

MS SILAT – Systèmes d'informations localisées pour l'aménagement des territoires



FUSION DE DONNEES LIDAR AEROPORTE ET SENTINEL-2 PAR APPRENTISSAGE POUR LA CARACTERISATION DE LA RESSOURCE FORESTIERE DU PNR DES BAUGES

Kamel LAHSSINI

Rapport de thèse professionnelle soutenue le 15 décembre 2021 à Montpellier



Sous la direction de **Sylvie DURRIEU** (INRAE)

The logo for INRAE, consisting of the letters "INRAE" in a bold, teal, sans-serif font.

TERRITOIRE ENVIRONNEMENT TELEDETECTION
INFORMATION SPATIALE

The logo for tetis, featuring a stylized globe icon with red, blue, and green segments, followed by the word "tetis" in a bold, dark blue, sans-serif font.



L'établissement AgroParisTech diffuse ce mémoire et se dégage de toute responsabilité sur le contenu proposé.

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition.

L'obligation de citation et de référencement lors de l'utilisation de ce document doit être respectée selon la licence Creative Commons apposée

Ce document est soumis aux conditions d'utilisation "Paternité – Pas d'utilisation commerciale – Pas de modification 4.0 France disponible en ligne : <https://creativecommons.org/licenses/by-nc-nd/4.0/>



SOMMAIRE

LISTE DES FIGURES	3
1 INTRODUCTION	4
1.1 Contexte.....	4
1.2 Etat de l'art	5
1.3 Démarche projet	7
1.3.1 <i>Objectif</i>	7
1.3.2 <i>Lots</i>	8
1.3.3 <i>Environnement et calendrier</i>	9
2 DE L'ETUDE DU TERRAIN.....	12
2.1 Présentation de la zone d'étude : le PNR des Bauges.....	12
2.2 L'inventaire forestier	14
2.3 Production d'une base de données terrain	16
3 DE LA TELEDETECTION LIDAR.....	20
3.1 Généralités sur le LiDAR aéroporté	20
3.2 Le LiDAR en foresterie	21
3.3 Données LiDAR Protest	22
4 DE LA CARTOGRAPHIE DE LA RESSOURCE FORESTIERE DU PNR DES BAUGES	25
4.1 Méthodologie.....	25
4.2 A partir de données LiDAR seules	26
4.3 Apport de la fusion de données multi-sources	30
4.4 Cartographies de la ressource forestière	34
5 CONCLUSION	37
6 ANNEXES.....	38
6.1 Base de données terrain	38
6.2 Métriques LiDAR.....	40
6.3 Architecture du modèle à 2 branches.....	41
6.4 Performances du modèle à 2 branches.....	41
7 BIBLIOGRAPHIE	43

LISTE DES FIGURES

Figure 1 - Découpage en lots et périmètre du projet	8
Figure 2 - Calendrier prévisionnel.....	10
Figure 3 - Localisation du PNR des Bauges	12
Figure 4 - PNR des Bauges (ESRI Shaded Relief ® et BD Forêt v2 ®)	13
Figure 5 - Vallée alpine du massif des Bauges en début de printemps © MEURET Michel – INRAE	13
Figure 6 - Une forêt mixte sur un flanc nord du massif des Bauges © MEURET Michel – INRAE	14
Figure 7 - Placettes d'inventaire du PNR des Bauges (ESRI Shaded Relief ®)	15
Figure 8 - Nuage de points 3D extrait de l'acquisition LiDAR sur le PNR des Bauges	20
Figure 9 – Processus de « rastérisation » du nuage © DAYAL Karun - INRAE	22
Figure 10 - « LiDAR Savoie zone B » (en violet) et « LiDAR zone PNR 74 » (en marron)	23
Figure 11 – Synthèse de la méthodologie.....	25
Figure 12 – Performances des modèles de prédiction à partir des données LiDAR seules	30
Figure 13 - Performances du modèle de prédiction par fusion en fonction des données d'entrée	33
Figure 14 - Cartographie de la surface terrière du PNR des Bauges (résolution 30 m).....	35
Figure 15 - Cartographie du volume total de bois du PNR des Bauges (résolution 30 m)	36
Figure 16 - Architecture du modèle à deux branches	41
Figure 17 - Surface terrière mesurée vs. estimée à partir du modèle multi-sources	42
Figure 18 - Volume total mesuré vs. estimé à partir du modèle multi-sources	42

1 INTRODUCTION

1.1 Contexte

De tout temps les écosystèmes forestiers ont été utilisés par l'Homme pour satisfaire ses besoins et contribuer au développement des sociétés. Pendant des siècles ces besoins ont majoritairement consisté en bois de chauffage, ce qui a conduit à une détérioration importante de la forêt française. Le développement de la marine et de l'industrie a aussi contribué à la surexploitation forestière. A la fin de la Révolution Française et face à la ruine programmée de ces espaces, les consciences se sont éveillées sur la nécessité de la mise en place d'une véritable gestion forestière et d'une sylviculture réfléchie.

La sylviculture consiste en l'ensemble des techniques permettant la création et l'exploitation rationnelle des forêts tout en assurant leur conservation et leur régénération. Les écosystèmes forestiers sont de ce point de vue des milieux certes naturels mais façonnés par l'Homme et les forestiers. D'après les données de l'IGN, la forêt en France métropolitaine couvre un peu plus de 30% du territoire et ce taux est en progression. La gestion de ces espaces représente donc un enjeu stratégique important, et ce d'autant plus que les objectifs visés se sont élargis au cours de ces cinquante dernières années : si la productivité reste un élément fondamental dans les stratégies de gestion, de nouvelles composantes sont néanmoins apparues à la faveur d'un contexte social et environnemental différent. La préservation de la biodiversité et la dimension sociale des territoires forestiers sont ainsi de nouvelles entrées à prendre en compte par les opérationnels et les décideurs dans leur politique de gestion.

Pour mener à bien ces missions et adopter la bonne stratégie en fonction du ou des objectifs visés pour le territoire, les gestionnaires ont besoin de s'appuyer sur des données de caractérisation et de description détaillées des écosystèmes forestiers. La connaissance précise et rationnelle du territoire d'étude va déterminer la nature et la temporalité des opérations sylvicoles adaptées à la stratégie choisie pour ce territoire. Les données de télédétection sont à ce titre utilisées pour mesurer des variables d'état et de structure de la végétation.

L'étude présentée dans ce rapport s'inscrit dans le cadre du projet de recherche Protest (PROspective TErritoriale SpaTialisée) financé par l'ADEME. Il s'agit d'un projet de recherche appliquée qui vise à construire une méthodologie d'analyse de la biomasse forestière en s'appuyant sur la technologie LiDAR mais également sur les avancées méthodologiques en modélisation technico-économique sous système d'information géographique, en simulation d'évolution forestière et en prospective territoriale. La télédétection n'est donc qu'une composante parmi d'autres dans la construction d'une

méthodologie plus globale qui mobilise diverses compétences et expertises. La zone d'étude retenue est le Parc Naturel Régional (PNR) du massif des Bauges. Le projet Protest s'articule autour de trois tâches principales. La première tâche consiste en la production d'une cartographie de la ressource forestière et de ses conditions de mobilisation. Elle se base notamment sur une analyse de données de télédétection LiDAR pour la cartographie des forêts. La deuxième tâche vise à mettre en place un travail de prospective territoriale et à définir des scénarios de gestion forestière qui combinent divers paramètres forestiers, environnementaux et sociétaux. Enfin dans la troisième tâche ces scénarios sont couplés avec un modèle d'évolution forestière spatialisé afin de simuler l'évolution du territoire en fonction de chaque scénario.

Le travail mené lors de cette mission professionnelle SILAT s'inscrit dans le cadre de la première tâche. La cartographie haute résolution de la ressource est essentielle car elle sert de base pour la modélisation de l'évolution forestière. Elle permet de spatialiser les résultats à différentes échelles et de mettre en évidence les hétérogénéités spatiales. En outre elle apporte une pertinence locale à des scénarios qui intègrent un contexte socio-économique plus large et forme ainsi un support particulièrement intéressant pour l'animation territoriale et les échanges entre les différents acteurs. La question de l'élaboration de cette cartographie est donc centrale et de premiers résultats ont déjà été produits par l'Office National des Forêts (ONF) à partir de relevés d'inventaire terrain et de données LiDAR seules. L'objectif du projet présenté dans ce rapport est d'évaluer le potentiel de la fusion de données issues de mesures LiDAR aéroporté (ALS) avec des images SENTINEL-2 pour la cartographie de la ressource forestière. La perspective de la fusion de données par méthodes d'apprentissage profond (deep learning) est d'améliorer la prédiction de certains attributs forestiers de structure estimés initialement à partir des données LiDAR uniquement.

1.2 Etat de l'art

Les écosystèmes forestiers jouent un rôle fondamental dans les équilibres naturels et dans les mécanismes climatiques par leur contribution au stockage du carbone. Plus de 40% du stock terrestre global de carbone est ainsi contenu dans ces écosystèmes (Pan et al. 2011; Beer et al. 2010). Ils abritent par ailleurs une biodiversité animale et végétale d'une grande richesse mais aussi très fragile (Simberloff 1999). Leur gestion durable et leur conservation sont donc des enjeux majeurs dans le contexte actuel de réchauffement climatique mondial.

En parallèle de ces considérations les espaces forestiers représentent un intérêt économique important grâce à la récolte et à l'exploitation du bois. Les ressources qui sont extraites contribuent au développement des sociétés et au maintien d'une filière bois qui implique de nombreux acteurs directs et indirects (Malfait et Martin 2017). Au cours de ces dernières décennies les changements dans les modes de vie des citoyens et leur perception de la nature ont aussi fait apparaître une dimension

sociétale des espaces forestiers qui est venue s'ajouter aux enjeux déjà existants de valorisation. Une nouvelle dimension économique est apparue par le développement du tourisme et des divertissements (Zhong, Deng, et Xiang 2008), ainsi qu'une dimension sociale liée à la qualité de vie et à la prise de conscience citoyenne sur les questions environnementales (DeFries et al. 2007; Simon et Bode 2012).

Pour répondre à ces problématiques et fournir aux gestionnaires les informations nécessaires à la prise de décisions, les données de télédétection sont devenues une source essentielle (Janssen 2004), et ce d'autant plus que l'offre s'est considérablement étoffée au cours de ces dernières décennies (Boyd et Danson 2005). Les capteurs déployés présentent une diversité de technologies utilisées et sont embarqués sur différents vecteurs aéroportés et satellitaires. L'information collectée couvre des zones plus ou moins étendues, parfois difficile d'accès voire inaccessibles, et permet d'accéder à des caractéristiques descriptives des environnements observés (Morgan, Gergel, et Coops 2010). La multiplication des données disponibles a entraîné par ailleurs le développement de nouvelles méthodologies pour leur traitement. Les données de télédétection sont aujourd'hui indispensables au suivi et à l'étude des écosystèmes naturels, et en particulier des forêts (Thenkabail 2016). Ces données en association avec des observations sur le terrain sont utilisées pour extraire des informations quantitatives sur les écosystèmes forestiers à différentes échelles de temps et d'espace, permettant ainsi un suivi efficace (DeFries et al. 2007), une meilleure description de leurs dynamiques (Cabello et al. 2012) et une gestion appropriée de la ressource forestière (Le Goff et al. 2010).

En particulier les informations relatives à la structure des forêts sont des descripteurs clés des écosystèmes forestiers (Franklin et al. 1981). Le terme structure fait référence à la répartition spatiale tridimensionnelle des éléments de la végétation tels que les troncs, les houppiers et les feuilles. Ces caractéristiques structurelles des écosystèmes forestiers sont directement liées à des grandeurs d'intérêts comme la quantité de biomasse (Chave et al. 2005). Les mesures précises de la structure de la forêt à partir de données de télédétection sont donc essentielles pour accéder à une meilleure connaissance de la ressource forestière. La technologie LiDAR est particulièrement adaptée pour ce type de mesures (Lim et al. 2003; van Leeuwen et Nieuwenhuis 2010). Utilisées avec des données de référence terrain, les données LiDAR permettent de développer des modèles empiriques pour prédire des paramètres forestiers à partir de variables de structure qui sont issues des nuages de points LiDAR. Une fois établis, les modèles sont ensuite extrapolés à la totalité de la zone d'étude pour établir la cartographie de paramètres forestiers (Bouvier et al. 2015; Véga et al. 2016).

Certaines études ont montré l'intérêt de combiner les données LiDAR avec l'information spectrale d'images provenant de capteurs spatiaux multispectraux à haute résolution spatiale (Monnet 2015; Wu et al. 2016) ou de capteurs hyperspectraux aéroportés (Luo et al. 2019). L'offre en données de télédétection spatiale adaptées au suivi des écosystèmes forestiers évolue rapidement. La mission

SENTINEL-2 apporte par exemple une richesse spectrale à haute résolution spatiale précieuse pour identifier la composition et l'état des peuplements (Karasiak et al. 2017; Grabska et al. 2019), et ce avec une couverture systématique et une fréquence de revisite inédite à ces résolutions. Certaines études ont également démontré un intérêt à combiner les informations radiométriques SENTINEL-2 avec les données LiDAR aéroportées pour l'estimation de la biomasse forestière (Wang et al. 2020). Du fait de leurs caractéristiques les données SENTINEL-2 apportent une information complémentaire à celle des données LiDAR et peuvent contribuer à améliorer l'estimation de variables forestières de structure (da Silveira et al. 2018).

En parallèle les approches d'apprentissage profond se sont développées et facilitent l'analyse de jeux de données de télédétection volumineux et complexes comme les séries temporelles (Interdonato et al. 2019), les données multi-sources (Ienco et al. 2019) ou les analyses multi-résolutions (Benedetti et al. 2018). Des études ont ainsi montré que les modèles par apprentissage profond pouvaient apporter une amélioration de la précision de l'estimation de la biomasse forestière par rapport à des méthodes plus communes, comme la régression linéaire multiple par exemple (García-Gutiérrez et al. 2016). De nouvelles approches de fusion de données optiques et LiDAR par deep learning ont également été développées récemment et ont permis de réaliser des cartographies de la ressource forestière à des échelles régionales (Narine, Popescu, et Malambo 2019; Zhang et al. 2019).

1.3 Démarche projet

1.3.1 Objectif

Comme spécifié dans la présentation du contexte cette mission professionnelle SILAT a pour but d'apporter une contribution à la composante cartographie haute résolution du projet Protest. En s'appuyant sur des méthodes d'apprentissage profond, l'objectif est d'évaluer le potentiel de la fusion de données issues de mesures LiDAR aéroporté (ALS) avec des images SENTINEL-2 pour la cartographie de la ressource forestière du PNR des Bauges. L'objectif principal vise à identifier les données pertinentes à dériver des nuages de points ALS et à définir une stratégie de fusion avec les données SENTINEL-2 pour améliorer la prédiction de certains attributs forestiers de structure estimés initialement à partir des seules données LiDAR. Les attributs forestiers en question sont principalement la surface terrière et le volume de bois.

Un travail de stage de fin d'études de M2 est également mené en parallèle sur l'utilisation des séries temporelles SENTINEL-2 pour la caractérisation des peuplements et leur association avec les données ALS pour améliorer la cartographie de la composition en essences. Le second objectif de cette mission professionnelle SILAT consiste donc en une contribution sur l'orientation et le suivi de ce travail de

stage dont les données et résultats intermédiaires seront par ailleurs utilisés pour répondre à l'objectif principal décrit plus haut.

1.3.2 Lots

La philosophie générale des démarches de cartographie consiste à associer des mesures terrain à des données de télédétection pour créer des modèles de prédiction de la variable que l'on souhaite représenter. Ces modèles peuvent ensuite être appliqués à l'ensemble de la zone d'étude pour produire les cartographies. La problématique de ce projet consiste à étudier le potentiel de la fusion de données multi-sources pour l'amélioration des résultats.

Ces quelques éléments font apparaître naturellement un découpage chronologique du projet en lots interconnectés par leurs entrées et leurs sorties :

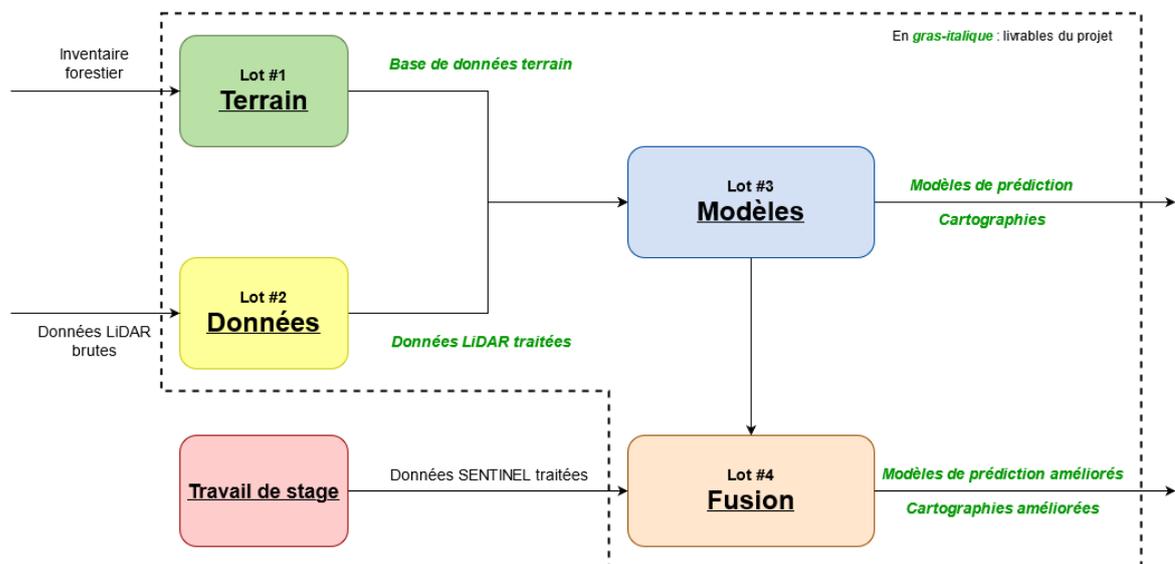


Figure 1 - Découpage en lots et périmètre du projet

L'avantage de ce découpage est qu'il permet de diviser le projet en fonction de livrables bien définis. Par ailleurs, bien qu'ils soient évidemment liés les uns aux autres, une partie des développements liés au contenu de chaque lot peut être menée de manière indépendante. Le périmètre de chacun des lots est défini comme suit :

- Lot « Terrain » (lot no. 1) : travaux liés à l'établissement d'une base de données (BDD) de référence terrain à partir des mesures d'inventaire forestier. Cette base de données sert de vérité terrain pour l'établissement des modèles et présente les variables dendrométriques de composition et de structure. Les livrables attendus sont une base de données correctement organisée et exploitable par d'autres utilisateurs, notamment grâce à un dictionnaire explicatif.

- Lot « Données » (lot no. 2) : travaux liés à la manipulation et aux traitements des données ALS. Les données brutes telles que livrées par le prestataire sont analysées pour identifier et produire des métriques LiDAR pertinentes dans l'établissement de modèles prédictifs. Les livrables attendus sont l'identification et le calcul de ces métriques d'intérêt, en veillant à nouveau à l'exploitabilité par d'autres utilisateurs.
- Lot « Modèles » (lot no. 3) : travaux liés au développement de modèles de prédiction par apprentissage à partir de données LiDAR ainsi qu'à leur implémentation pour la réalisation de cartographies. Les livrables des deux lots précédents constituent les entrées indispensables aux activités de ce troisième lot et les modèles développés doivent pouvoir s'adapter facilement à des entrées susceptibles d'évoluer : par exemple une mise à jour de la base de données terrain (sortie du lot no. 1) ne doit affecter ni la conception ni le fonctionnement des méthodes développées dans le cadre de ce lot. Les livrables attendus sont des modèles calibrés et des cartographies de la ressource forestière.
- Lot « Fusion » (lot no. 4) : travaux liés à l'intégration des données SENTINEL-2. La perspective de ce projet est d'étudier l'apport des données radiométriques dans les approches prédictives du lot « Modèles » (lot no. 3). Les données SENTINEL-2 traitées et exploitables sont fournies grâce à un travail de stage de M2. Réciproquement les livrables des lots no.1 et 2 de ce projet sont fournis au stagiaire pour son travail. Les livrables attendus sont des modèles et des cartographies de la ressource forestière améliorés par rapport aux approches à partir de données LiDAR seules.

Ce découpage fonctionnel permet de bien appréhender la stratégie mise en place pour répondre à l'objectif du projet et présente l'avantage de diviser le travail en lots indépendants, permettant ainsi un avancement en parallèle sur chacun des lots plutôt qu'un enchaînement chronologique linéaire très dépendant des livrables intermédiaires.

1.3.3 Environnement et calendrier

La mission se déroule dans les locaux de la Maison de la Télédétection à Montpellier, au sein de l'INRAE et sous la direction de Sylvie DURRIEU. Compte tenu du découpage du projet présenté plus haut, et notamment de la présence de livrables intermédiaires nécessaires au bon déroulé du projet, le calendrier a été établi pour prendre en compte la possibilité de revenir sur ces livrables si le besoin s'en faisait ressentir. En effet, l'avancement dans les travaux peut conduire à l'apparition de nouvelles hypothèses ou de nouveaux enjeux et ce calendrier permet de prendre en compte de tels changements.

Lot	Jalon	Démarrage	Echéance
Tous	Etude bibliographique	T0	T0 + 3 semaines
1	Première version consolidée de la BDD terrain	T0 + 3 semaines	T0 + 2 mois
2	Prise en main pratique et théorique des données LiDAR : formation aux notions et outils	T0 + 3 semaines	T0 + 2 mois
4	Suivi du travail de stage de M2 sur les données SENTINEL	T0 + 3 semaines	T0 + 5 mois
1	Version finalisée de la BDD suite à la validation des hypothèses et au RETEX utilisateurs	T0 + 2 mois	T0 + 3 mois
2	Première production de métriques LiDAR	T0 + 2 mois	T0 + 3 mois
3	Implémentation et validation de modèles d'apprentissage	T0 + 3 mois	T0 + 5 mois
1 ; 2	Reprise éventuelle des livrables intermédiaires lots 1 et 2 en parallèle de l'établissement des modèles d'apprentissage	T0 + 3 mois	T0 + 5 mois
4	Intégration des données optiques SENTINEL	T0 + 5 mois	T0 + 6 mois
Tous	Production des résultats de cartographie et restitution Rédaction du rapport et soumission d'un article scientifique	T0 + 6 mois	T0 + 8 mois

Figure 2 - Calendrier prévisionnel

Par ailleurs l'enchaînement logique et linéaire des jalons présentés ici implique une certaine dépendance entre les lots, les sorties des uns constituant les entrées des autres et vice versa. Pour autant il est aussi indispensable que chaque lot puisse être traité de manière indépendante : une attention particulière est donc portée à la conception et au formalisme des entrées et sorties pour permettre une flexibilité dans la production des livrables et garantir dans le même temps la modularité du projet. Par exemple une mise à jour de la BDD terrain (sortie du lot no. 1 et entrée du lot no. 3) se doit d'être la plus transparente possible pour les modèles d'apprentissage établis dans le cadre du lot no. 3 et ne doit pas occasionner de travail majeur lié à cette évolution ou à un changement de conception. Chaque lot doit donc pouvoir être traité comme un sous-projet indépendant et la démarche de conduite de projet présentée en figure 1 répond à ce besoin.

Enfin le T0 évoqué en figure 2 correspond à la date de fin des cours MS SILAT et au début de la période de mission professionnelle à temps plein. Il est important de souligner que cette mission a été effectuée en apprentissage : durant la période académique qui a précédé la période de mission professionnelle, des semaines d'intégration en entreprise intercalées entre les cours ont permis de se

familiariser avec le fonctionnement de l'Unité Mixte de Recherche (UMR) d'accueil et d'intervenir sur des missions ponctuelles en support à d'autres travaux menés par des étudiants sur des thématiques proches. Ces travaux ont ainsi offert l'opportunité d'une première prise en main des données et outils associés au projet. Ils ont aussi été l'occasion d'échanges à la fois très formateurs pour la suite de la mission mais aussi très utiles pour l'intégration au sein de l'UMR. En particulier, la participation à une campagne de mesures terrain en forêt sur le site du Ciron dans les Landes a été une formidable expérience humaine et professionnelle. Cette période d'alternance a donc été formatrice à bien des égards et a permis d'entamer la mission professionnelle avec un certain capital de départ.

2 DE L'ETUDE DU TERRAIN

2.1 Présentation de la zone d'étude : le PNR des Bauges

La finalité du projet Protest est de produire une méthodologie générale de prospective territoriale spatialisée. A terme, l'idée serait donc de pouvoir appliquer les principes généraux de cette méthodologie à n'importe quel territoire. Le test de cette méthodologie est réalisé sur le site pilote du Parc Naturel Régional du Massif des Bauges (figure 3). Ce territoire constitue la zone d'étude du projet.



Figure 3 - Localisation du PNR des Bauges

En France, un Parc Naturel Régional (PNR) est un territoire ayant choisi volontairement d'adopter un mode de développement basé sur la mise en valeur et la protection de ses patrimoines naturels et culturels. Situé à cheval entre les départements de Savoie (73) et de Haute-Savoie (74), le massif des Bauges est labellisé « géoparc » et est reconnu au niveau national pour sa forte valeur patrimoniale et paysagère. Sa superficie forestière est de 51 136 hectares et représente environ 60% de la superficie totale du PNR (figure 4). La forêt est donc un élément structurant du massif, tant sur le plan économique (ressource en bois, tourisme, ...) que sur le plan environnemental (biodiversité, habitats d'espèces, trame verte, ...). Des problématiques différentes coexistent sur ce territoire avec d'une part la production de bois (75% du massif est classé en forêt de production) et d'autre part la conservation du patrimoine (41% de la surface totale du PNR est concernée par un zonage révélant un intérêt patrimonial).

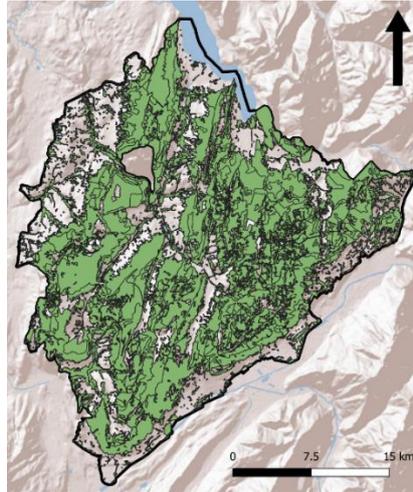


Figure 4 - PNR des Bauges (ESRI Shaded Relief ® et BD Forêt v2 ®)

Le territoire se caractérise par une forte hétérogénéité à plusieurs niveaux.

- Tout d’abord sur le plan du relief : situé dans la sylvoécocorégion (SER) des Préalpes du Nord, le géoparc présente de fait une topographie irrégulière et montagneuse, alternant plaines et sommets. Le relief a un impact important sur les données de télédétection et rend aussi les campagnes de relevés terrain plus complexes.
- Ensuite sur le plan de l’occupation du territoire : si la forêt couvre la majorité de la superficie de la zone elle coexiste avec d’autres types d’espaces, notamment des prairies non boisées et des espaces agricoles, sans compter les aménagements urbains liés aux fonctions sociétales et économiques du géoparc. Un soin particulier devra donc être apporté à l’analyse des données et aux résultats de cartographie dans la mesure où cette étude se concentre sur l’estimation de la ressource forestière.
- Enfin sur le plan des peuplements forestiers : les espaces forestiers présentent une grande diversité d’essences avec une coexistence d’espèces feuillues et résineuses. La caractérisation de la composition spécifique, qui entre dans le cadre du travail de stage de M2 évoqué précédemment, est un enjeu important de la description de la ressource forestière du PNR.



Figure 5 - Vallée alpine du massif des Bauges en début de printemps © MEURET Michel – INRAE



Figure 6 - Une forêt mixte sur un flanc nord du massif des Bauges © MEURET Michel – INRAE

Cette hétérogénéité dans l'état et la structure du territoire, illustrée par les figures 5 et 6, introduit une complexité supplémentaire dans la démarche du projet Protest et doit être prise en compte dans la démarche de cartographie de la ressource forestière présentée dans ce rapport.

2.2 L'inventaire forestier

La création de modèles empiriques de prédiction d'attributs forestiers nécessite d'associer des données de télédétection avec des mesures ciblées sur le terrain de ces mêmes attributs. La mesure précise des variables dendrométriques d'intérêt sur le terrain représente donc un enjeu majeur pour la pertinence des modèles établis et l'interprétation de leurs résultats. Par ailleurs il est en général complexe de mesurer directement sur le terrain la valeur des attributs qu'on souhaite cartographier, que ce soit pour des raisons pratiques (temps, moyens, ...) ou des raisons liées à la nature de l'attribut mesuré (par exemple le volume de bois ne peut être mesuré directement). Un protocole d'inventaire forestier a été mis en place dans le cadre du projet Protest pour effectuer un certain nombre de relevés et de mesures qui permettent ensuite d'établir la vérité terrain qui sert d'entrée à la création des modèles de prédiction (livrable du lot no. 1). Il est important de noter dès à présent que cette vérité terrain est estimée à partir de mesures d'inventaire partielles et peut donc à ce titre présenter des biais compte tenu du fait qu'elle dépend des différentes hypothèses prises pour l'estimation. Ce point très important est discuté à maintes reprises dans la suite de ce rapport.

De manière générale un inventaire forestier consiste à évaluer la ressource d'une forêt et à décrire précisément un peuplement. Des paramètres tels que le diamètre des arbres et leur espèce sont relevés, ainsi que d'autres informations qui permettent de décrire les peuplements, comme le type de sol par exemple. Les mesures effectuées sont en général géoréférencées afin de pouvoir les associer correctement avec les données de télédétection. Les méthodes d'inventaire se divisent en deux catégories en fonction de la zone à décrire : les inventaires pied par pied, où tous les arbres sont inventoriés, et les inventaires statistiques, où seule une partie de la forêt est inventoriée. Dans le cas d'un inventaire pied par pied tous les arbres de la forêt sont identifiés et recensés. En théorie cette

méthode donne une connaissance exacte du peuplement (tous les individus sont reconnus et mesurés). Mais en pratique elle est très coûteuse voire impossible à mettre en place, en particulier sur une zone d'étude étendue et hétérogène comme le PNR des Bauges. C'est donc la seconde méthode qui a été choisie pour l'inventaire du projet Protest. L'inventaire statistique consiste à choisir un certain nombre de placettes de surface déterminée réparties sur l'ensemble du territoire et à y effectuer les mesures de paramètres forestiers. Le nombre des placettes ainsi que leur maillage sont choisis en fonction de la variabilité du peuplement : une grande forêt très homogène nécessite peu de placettes d'inventaire malgré sa surface étendue tandis qu'un petit massif très varié ne peut être correctement décrit qu'à partir d'un nombre suffisant de placettes en dépit de sa taille réduite. La philosophie de ce type d'inventaire est donc de pouvoir tirer des conclusions à l'échelle d'un territoire entier à partir de relevés localisés mais suffisamment bien ciblés pour couvrir toutes les caractéristiques des peuplements présents.

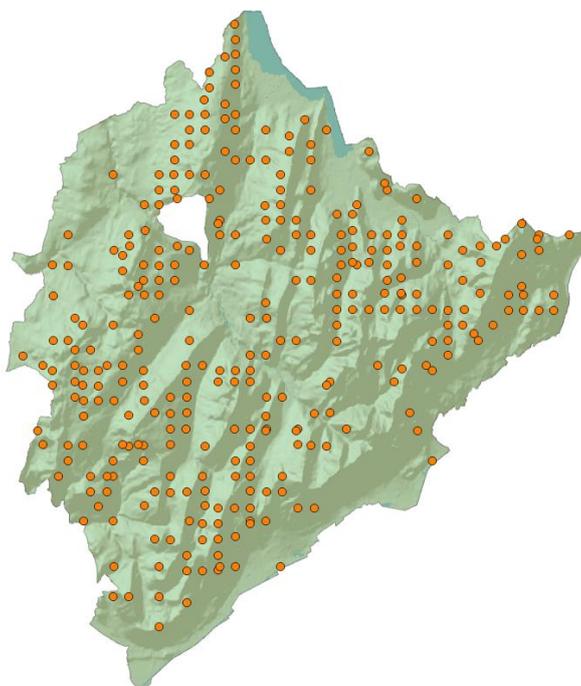


Figure 7 - Placettes d'inventaire du PNR des Bauges (ESRI Shaded Relief ®)

L'inventaire forestier du projet Protest a été réalisé conjointement par l'agence Savoie de l'ONF et par l'INRAE en septembre 2018. 291 placettes régulièrement réparties sur le PNR des Bauges ont été visitées par les techniciens et plusieurs paramètres forestiers y ont été mesurés. Deux classes d'arbres ont été choisies pour définir la stratégie d'inventaire : la classe 10, qui correspond aux arbres dont le diamètre à hauteur de poitrine (DBH) est compris entre 7,5 cm et 17,5 cm, et la classe 15, qui correspond aux arbres dont le DBH est supérieur à 17,5 cm. Les mesures d'inventaire ont été réalisées sur des placettes circulaires concentriques de 10 m et 15 m de rayon, d'où le libellé des classes. Les

arbres de la classe 15 ont été précisément relevés dans la placette de 15 m : leur DBH, leur espèce et leur position dans la placette ont ainsi été mesurés par les techniciens. Concernant les arbres de la classe 10 ils ont été simplement comptés dans la placette de 10 m : cela signifie qu'aucune information individuelle relative à leur DBH ou leur essence n'est disponible dans les mesures d'inventaire. Par ailleurs la hauteur des arbres n'a été mesurée pour aucune des classes d'inventaire. Or ces données sont indispensables à l'estimation à l'échelle de la placette d'attributs forestiers tels que la surface terrière ou le volume de bois. Par conséquent les seules mesures d'inventaire du projet Protest sont insuffisantes à l'établissement d'une vérité terrain exhaustive et cette dernière devra donc être estimée en prenant un certain nombre d'hypothèses pour combler les informations manquantes. Ce protocole d'inventaire particulier tient au fait que la variable généralement ciblée par les forestiers est la surface terrière des arbres de DBH supérieur à 17,5 cm. Cependant les arbres de DBH inférieur contribuent aussi à la mesure LiDAR par leur présence dans les nuages de points. Il est donc nécessaire de les prendre en compte pour assurer une cohérence entre la vérité terrain et les données de télédétection.

2.3 Production d'une base de données terrain

L'objectif de la production d'une base de données terrain est de calculer les trois variables forestières suivantes pour chaque placette d'inventaire : la surface terrière, le volume total de bois et le Gini des diamètres. La surface terrière (notée G) est un indice très utilisé par les forestiers et correspond, pour un arbre donné, à la surface de la section de l'arbre mesurée à hauteur de poitrine, soit à 1,30 mètre du sol. La surface terrière d'une placette d'inventaire est la somme des surfaces terrières de tous les arbres qu'elle contient et cette valeur est ensuite normalisée par rapport à la surface de la placette. Elle s'exprime en m²/ha (mètres carrés par hectare). Le volume total de bois (noté V) est aussi un indicateur très important de la ressource forestière. Un arbre est composé d'une tige (le tronc) et d'un houppier. L'ensemble des branches et des feuilles constitue le houppier. Le volume total de bois d'un arbre inclut tout le bois contenu dans l'individu, à savoir la tige et les branches. Il est calculé à partir de tables prédéfinies appelées tarifs de cubage et qui prennent en entrée le DBH, la hauteur et l'espèce de l'arbre. De la même manière que pour le G, le volume total d'une placette d'inventaire est la somme des volumes totaux de tous les arbres qu'elle contient. Il s'exprime en m³/ha (mètres cubes par hectare). Enfin le coefficient de Gini est une mesure statistique qui renseigne sur l'hétérogénéité d'une distribution. Il est calculé sur la distribution des DBH des arbres de la placette et mesure donc l'hétérogénéité du peuplement. Il est compris entre 0 et 1 (0 pour une distribution uniforme ; 1 pour une distribution totalement hétérogène) et n'a pas d'unité.

L'objet placette constitue l'unité de base de l'étude présentée ici : toutes les variables sont calculées à l'échelle de la placette et les modèles de prédiction sont calibrés à partir de données calculées à cette échelle (variables forestières, métriques LiDAR, données exogènes). Comme évoqué dans la partie précédente les relevés d'inventaire ont été réalisés sur des placettes circulaires concentriques de rayons

10 m et 15 m. Par conséquent l'échelle retenue est celle de la placette la plus grande, à savoir 15 m de rayon. La base de données terrain est donc constituée de 291 objets placette de 15 m de rayon pour lesquels sont estimées les trois variables forestières d'intérêt. Le calcul de ces trois variables nécessite de connaître pour chacun des arbres contenus dans les placettes de 15 m les informations suivantes : le DBH (pour la surface terrière et le Gini) ainsi que la hauteur totale et l'espèce d'appartenance (pour le volume total). Or le protocole d'inventaire n'a pas permis de couvrir l'ensemble de ces besoins. D'une part les arbres de la classe 10 n'ont pas été mesurés individuellement (leur DBH, leur hauteur totale et leur espèce ne sont pas connus) et ils n'ont été comptés que dans une placette de 10 m de rayon. D'autre part la hauteur totale n'a été mesurée pour aucune des deux classes d'arbre (10 et 15). Il est donc nécessaire de combler ces manques en ajoutant de l'information.

Dans un premier temps, les informations relatives à la placette de 10 m ont été ramenées à l'échelle de 15 m par l'application d'une relation proportionnelle de coefficient égal au rapport de la surface de la placette de 15 m par celle de 10 m. L'hypothèse d'une distribution spatiale homogène du peuplement permet d'estimer par cette méthode le nombre d'arbre de la classe 10 dans une placette de 15 m avec une erreur théoriquement acceptable mais non vérifiable d'un point de vue pratique. Le nombre d'individus de la classe 10 dans chaque placette de 15 m est donc une première approximation.

Ensuite la question ouverte principale est celle de la détermination des diamètres des arbres de la classe 10. En effet le DBH intervient dans le calcul des trois variables forestières à prédire et il n'a pas été mesuré pour les arbres de cette classe. Par conséquent cette information doit être ajoutée artificiellement en prenant l'hypothèse qui minimise l'erreur par rapport à la réalité du terrain. Pour ce faire deux hypothèses ont été testées : affecter à tous les arbres de la classe 10 une même valeur de DBH constante (par exemple la médiane de la classe, i.e. 12,5 cm, ou une autre valeur à déterminer) ou bien pour chaque individu une valeur aléatoire tirée dans l'intervalle [7,5 ; 17,5]. Pour évaluer l'erreur introduite par l'une ou l'autre des hypothèses les données de l'inventaire forestier national (IFN) produit par l'IGN ont été utilisées. Dans cette base de données externe, qui couvre l'ensemble du territoire français, tous les arbres de DBH supérieur à 7,5 cm ont été individuellement mesurés. Il est donc possible à partir de ces données d'inventaire disponibles gratuitement en ligne de calculer directement une surface terrière et un coefficient de Gini des diamètres. Les placettes d'inventaire de l'IFN de 2008 à 2019 et localisées au niveau de la zone du PNR des Bauges ont donc été extraites pour étudier l'impact des hypothèses mentionnées plus haut : d'une part les attributs forestiers « réels » sont calculés à partir des mesures de l'IFN et d'autre part les hypothèses sur les valeurs de DBH sont appliquées aux placettes de l'IFN (i.e. les valeurs de DBH réellement mesurées sont remplacées soit par une valeur constante, soit par des valeurs tirées aléatoirement) pour recalculer ces mêmes attributs forestiers « estimés ». L'hypothèse qui minimise l'erreur entre la réalité et l'estimation est celle retenue pour combler les informations de DBH manquantes dans l'inventaire forestier Protest. Par

ailleurs l'étude de la distribution des arbres des placettes de l'IFN dont le DBH est compris entre 7,5 et 17,5 cm permet de raffiner les hypothèses testées : il en ressort que la distribution de DBH est presque uniforme dans l'intervalle [7,5 ; 17,5] et que la valeur médiane est égale à 11,1 cm, soit légèrement inférieure au centre de la classe. Par conséquent les deux hypothèses testées pour combler l'inventaire Protest sont précisées : soit affecter à tous les arbres de la classe 10 une valeur constante égale à 11,1 cm (hypothèse no. 1), soit effectuer pour chaque individu un tirage aléatoire selon une loi uniforme dans l'intervalle [7,5 ; 17,5] (hypothèse no. 2). Ainsi la surface terrière et le coefficient de Gini des placettes IFN sont calculés tour à tour selon ces deux hypothèses et comparés ensuite aux valeurs de référence obtenues avec les véritables données de l'IFN. Cette étude permet à la fois de privilégier la bonne hypothèse mais aussi de quantifier l'erreur associée. Le premier résultat d'intérêt est que l'hypothèse no. 1 n'introduit pas de biais fort dans le calcul de la surface terrière : cela signifie que le DBH des arbres de la classe 10 de l'inventaire Protest peut être fixé à 11,1 cm pour le calcul de cet attribut forestier. Le second résultat à noter est que l'hypothèse no. 1 est moins adaptée au calcul du coefficient de Gini pour les placettes qui contiennent une majorité d'arbres de classe 10. En effet, par définition, ce coefficient mesure l'hétérogénéité d'une distribution et par conséquent l'affectation d'une même valeur fixe de DBH a tendance à créer artificiellement de l'homogénéité dans une distribution qui peut être plus ou moins hétérogène. A l'inverse, on constate pour ces placettes que l'hypothèse no. 2 conserve l'hétérogénéité relative et introduit une erreur acceptable dans le calcul de cet attribut forestier. On note également que pour les placettes où les arbres de classe 10 sont moins nombreux l'hypothèse no. 1 donne de meilleurs résultats pour le calcul du coefficient de Gini. Ces observations permettent d'orienter précisément les choix pour compléter l'inventaire Protest. L'hypothèse no. 1 est donc celle retenue pour le calcul de la surface terrière. Pour le Gini des DBH une approche hybride est adoptée : pour les placettes Protest dont la proportion d'arbres de classe 10 est inférieure à 65% c'est l'hypothèse no. 1 qui est appliquée ; pour les autres c'est l'hypothèse no. 2 qui est la plus adaptée.

Concernant le volume total, il est calculé à partir de tarifs dits « à deux entrées ». Pour un arbre donné ces tarifs calibrés prennent en entrée son DBH et sa hauteur totale et dépendent aussi de son espèce d'appartenance. Or comme précisé plus haut la hauteur totale des arbres n'a pas été mesurée, que ce soit pour les arbres de la classe 10 ou ceux de la classe 15. Là encore cette information doit donc être ajoutée pour pouvoir appliquer les tarifs aux arbres des placettes et estimer le volume de bois total. Pour ce faire, il existe des relations allométriques qui lient la hauteur d'un arbre à son DBH en fonction de l'espèce étudiée. A partir de ces relations et de la connaissance du DBH (soit mesuré pour les arbres de la classe 15, soit estimé pour les arbres de la classe 10) il est alors possible de calculer une valeur de hauteur. L'enjeu est donc celui de la construction de ces relations allométriques et, là encore, la base de données de l'IFN fournit les données nécessaires. Le PNR des Bauges est localisé dans la sylvoécocorégion des Préalpes du Nord (SER H10). Les sylvoécocorégions sont un découpage

géographique des écosystèmes forestiers en France et présentent une cohérence en termes de climat et de caractéristiques écologiques (sols, ...). Pour la construction de relations DBH-hauteur il est donc pertinent de considérer les individus de l'ensemble de la SER et pas uniquement ceux localisés au niveau du PNR comme dans la démarche précédente. Les placettes d'inventaire de l'IFN de 2008 à 2019 et localisées au niveau de la SER H10 ont donc été extraites et, pour chaque essence identifiée, les nuages de points de la hauteur en fonction du DBH ont été tracés. Le fait de considérer tous les individus de la SER permet de disposer d'un plus grand nombre de couples [DBH, hauteur] pour la création des relations allométriques. Pour chaque essence identifiée, un modèle puissance de la forme $hauteur = a \cdot DBH^b$ est calé sur les données de l'IFN, avec a et b les paramètres du modèle. Grâce à la bibliothèque de modèles ainsi créée la hauteur de chaque arbre de l'inventaire Protest peut être estimée en appliquant le modèle correspondant à l'espèce de l'individu. Une fois les hauteurs calculées les tarifs « à deux entrées » peuvent être appliqués à tous les individus de l'inventaire Protest pour estimer le volume de bois à l'échelle de la placette. Concernant les arbres de la classe 10 dont le DBH et l'essence sont inconnus une stratégie particulière a été mise en place. L'essence précise de ces arbres est certes inconnue mais la famille d'appartenance (arbre feuillu ou résineux) a bien été relevée sur le terrain. Des relations allométriques globales par grande famille ont donc également été dérivées des données de l'IFN puis utilisées pour estimer les hauteurs des arbres de la classe 10. Par ailleurs les deux hypothèses de l'étude précédente ont également été testées de la même manière pour voir comment elles affectent l'estimation du volume total au niveau des placettes. Il en ressort une conclusion analogue à celle observée pour la surface terrière, à savoir que l'hypothèse no.1 est la plus adaptée pour le calcul du volume de bois.

Les travaux présentés dans cette partie se sont inscrits dans le périmètre du lot no.1 et ont permis de produire un premier livrable important : une base de données terrain qui contient plusieurs variables forestières d'intérêt à l'échelle de la placette, et notamment les trois attributs de structure que l'on souhaite cartographier sur toute la zone du PNR des Bauges. De nombreuses autres variables forestières, d'état comme de structure, ont également été produites dans le cadre de l'élaboration de cette BDD et ces travaux ne sont pas décrits de manière exhaustive ici (cf. annexe 6.1). Comme cela a été présenté, il est important de garder à l'esprit que cette vérité terrain a été établie selon certaines hypothèses dont l'impact a été étudié, et que par conséquent les résultats discutés plus bas dans ce rapport se doivent d'être analysés en prenant en compte cet aspect-là.

3 DE LA TELEDETECTION LIDAR

3.1 Généralités sur le LiDAR aéroporté

L'acronyme LiDAR (Light Detection and Ranging) désigne une technique de télédétection active qui repose sur l'émission et la réception d'une impulsion laser. Elle permet d'obtenir la distance entre le récepteur et la cible à partir de la vitesse de déplacement de l'impulsion laser (célérité de la lumière) et de son temps de parcours (temps écoulé entre l'émission de l'impulsion et la réception d'un écho).

Dans le cadre de ce projet les données LiDAR ont été acquises de manière aéroportée : un avion muni d'un système de positionnement tridimensionnel (GPS pour la position et centrale inertielle pour l'attitude) embarque un capteur LiDAR, équipé d'un mécanisme de déviation du faisceau laser (mécanisme à balayage). Le capteur émet une impulsion laser, en général dans les longueurs d'onde du proche infra-rouge, qui va traverser l'atmosphère et interagir avec les cibles au sol. Cette interaction induit un signal retour par rétrodiffusion qui va être enregistré par le capteur et traité numériquement. La forme du signal retour dépend de plusieurs paramètres comme les propriétés radiométriques et géométriques des objets, l'environnement de l'acquisition ou encore le diamètre du faisceau laser. Le traitement numérique de ce signal peut aboutir en sortie soit à un ensemble d'impulsions retour (cas du LiDAR à retour discret), soit à un signal continu (LiDAR à forme d'onde complète). Les données utilisées dans le cadre de ce projet ont été obtenues à l'aide d'un capteur à retour discret. Le signal est alors traité pour fournir un nuage de points 3D géoréférencés grâce à la connaissance du positionnement tridimensionnel du capteur lors de l'acquisition. Chaque point de ce nuage correspond à un écho lié à l'interaction entre l'impulsion laser et l'environnement. On distingue les échos simples (retour unique) des échos multiples (plusieurs rétrodiffusions enregistrées pour une même impulsion laser).

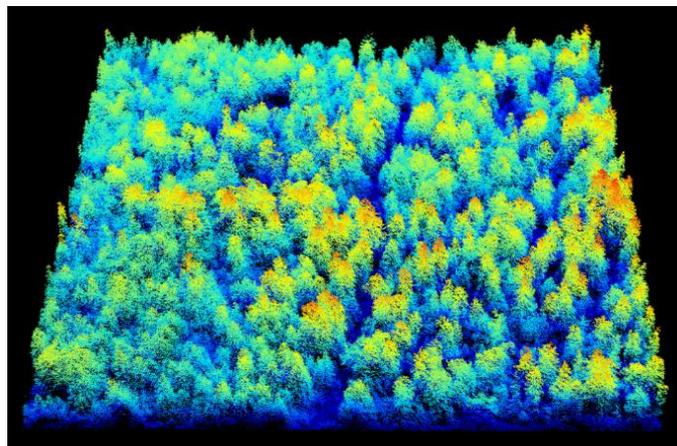


Figure 8 - Nuage de points 3D extrait de l'acquisition LiDAR sur le PNR des Bauges

La figure 8 permet d'appréhender le type et la forme des données issues d'une acquisition LiDAR à retour discret. La structure tridimensionnelle de l'environnement observé par le capteur apparaît visuellement sur le nuage de points : couronnes des arbres, végétation, troncs, ... De manière générale le nuage de points peut être caractérisé par une série de métriques décrivant sa distribution dans l'espace et certaines propriétés du signal. Il est également possible d'en dériver plusieurs produits raster : métriques rastérisées, modèle numérique de terrain (MNT), modèle numérique de surface (MNS), ... Le format standardisé de distribution du nuage de points LiDAR est le format LAS (LASer file format exchange), dont les spécifications sont officiellement définies et qui contient toutes les informations permettant de caractériser le nuage de points : coordonnées des points, intensité associée, numéro d'écho, angle de scan, ... La plupart des traitements présentés dans ce rapport ont été réalisés en langage de programmation R grâce à l'excellent package « lidR », qui permet de manipuler et de visualiser des données LiDAR au format LAS.

3.2 Le LiDAR en foresterie

Les peuplements forestiers sont décrits selon deux grandes familles d'attributs : les paramètres de composition d'une part, qui renseignent sur les espèces présentes et la biodiversité, et les paramètres de structure d'autre part, qui décrivent les peuplements de manière structurelle (hauteurs, surfaces, volumes, ...) et biophysique. De par le principe de la technologie et le format des données, les mesures LiDAR sont particulièrement adaptées à l'estimation des paramètres de cette seconde famille. En effet comme vu dans la partie précédente la particularité de la technique de télédétection LiDAR est qu'elle permet d'obtenir une représentation tridimensionnelle explicite de la zone observée. Elle est donc très bien adaptée à la description structurelle et géométrique de la forêt. Ce type d'information est essentiel pour estimer la ressource forestière et il existe un large éventail de méthodes basées sur l'analyse de données LiDAR pour une caractérisation améliorée des écosystèmes forestiers et une meilleure compréhension de leur fonctionnement.

Ces méthodes peuvent être rangées selon deux niveaux principaux en fonction du type d'informations recherchées sur un peuplement donné : le niveau de la « placette » et celui de « l'arbre ». L'approche par « arbre » consiste à segmenter le nuage de points LiDAR afin d'isoler les individus et de reconnaître les points correspondant à un arbre spécifique. Une fois la segmentation effectuée, des variables descriptives peuvent être extraites du nuage de points LiDAR associé à un objet (i.e. un individu) et fournir une information à un niveau de détail relativement fin. Ces méthodes sont néanmoins très dépendantes de la qualité de la segmentation, qui peut être très variable selon les peuplements, et les temps de calcul associés peuvent être longs.

Dans l'approche « placette », qui est celle adoptée dans le cadre de ce travail, des modèles empiriques sont calibrés à l'échelle de la placette en liant la variable dendrométrique d'intérêt relevée sur le

terrain aux données LiDAR extraites à cette échelle. De manière générale, les nuages de points 3D fournissent beaucoup d'informations mais ils sont difficilement exploitables en tant que tels, non seulement du fait de la taille des données mais aussi de leur formalisme particulier. Des métriques LiDAR sont donc dérivées des nuages de points à partir d'une grille de résolution donnée et permettent de représenter l'information sous un format différent : on parle alors d'approche surfacique dans la mesure où l'information 3D est ramenée à l'échelle de la placette ou du pixel (cf. figure 9). La résolution de la grille est choisie de telle manière que la surface d'un pixel soit équivalente à celle de la placette.

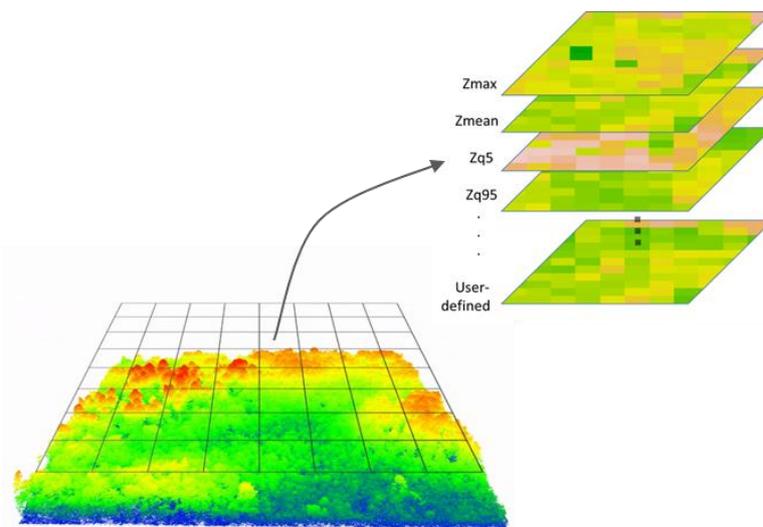


Figure 9 – Processus de « rasterisation » du nuage © DAYAL Karun - INRAE

La rasterisation du nuage de points consiste pour chaque pixel de la grille à extraire le sous-nuage contenu dans l'emprise du pixel et à calculer un certain nombre de métriques descriptives de ce sous-ensemble de données. On passe alors d'un format nuage de points tridimensionnel à une pile de rasters dont chaque bande correspond à une métrique donnée. Les modèles calibrés peuvent alors être appliqués à chaque pixel de la grille pour prédire l'attribut dendrométrique étudié. Cette démarche, qui constitue le fil directeur de ce travail, est explicitée tout au long de la suite de ce document.

3.3 Données LiDAR Protest

Dans la partie précédente l'importance de disposer d'une vérité terrain fiable et conforme à la réalité a été soulignée. Les données de télédétection constituent la seconde entrée des modèles empiriques et sont à ce titre tout aussi déterminantes dans la prédiction des attributs forestiers.

Les données LiDAR utilisées dans le cadre du projet Protest ont été obtenues lors de deux campagnes d'acquisition qui ont permis de couvrir l'ensemble du PNR des Bauges. Il est important de noter que les opérations d'acquisition de données sont réalisées en période « en feuilles ». Ces campagnes

diffèrent par la zone couverte, le matériel utilisé et la date de leur réalisation. Ainsi le jeu de données est divisé selon deux sources :

- « LiDAR Savoie zone B » : l'acquisition a été réalisée par l'IGN entre juin et septembre 2016 avec une densité d'émission moyenne de 4 points/m². Elle couvre une surface d'environ 53 600 hectares. Le traitement des données, qui inclut la création des produits dérivés tels que le MNT et la correction des nuages de points en hauteur, a été réalisé par la société Sintégra. La zone couverte par l'acquisition correspond environ à la partie du PNR située dans le département 73 et elle contient 136 placettes terrain.
- « LiDAR zone PNR 74 » : l'acquisition a été financée par le PNR des Bauges dans le cadre du projet Protest afin de compléter les secteurs non couverts par les données « LiDAR Savoie zone B ». Elle a été réalisée par la société Opsia en septembre 2018 avec une densité d'émission moyenne de 13 points/m². La densité d'émission présente une répartition géographique hétérogène (comprise en 4 et 25 points/m²) en raison des variations de terrain liées à la topographie montagneuse. La zone décrite par l'acquisition couvre une surface d'environ 37 350 hectares et correspond à la partie du PNR située dans le département 74. Elle contient 180 placettes terrain.

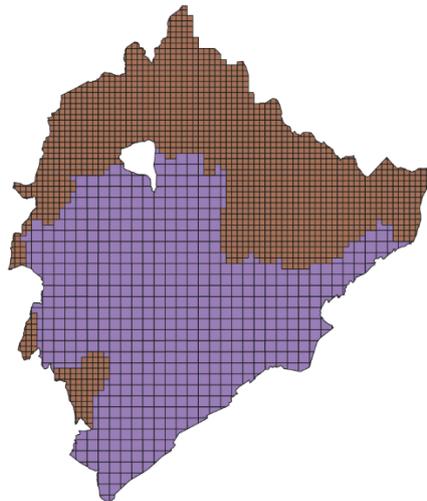


Figure 10 - « LiDAR Savoie zone B » (en violet) et « LiDAR zone PNR 74 » (en marron)

La différence de spécifications entre les missions LiDAR, que ce soit en termes techniques (densité d'émission) ou pratiques (période d'acquisition), apparaît donc assez importante. Néanmoins, compte tenu de la méthodologie présentée ici et qui consiste à calculer des métriques à l'échelle de la placette, les données LiDAR issues de ces deux acquisitions peuvent être considérées comme un jeu de données unique et les deux zones n'ont pas à être traitées de façon indépendante. En effet il a d'abord été vérifié que la différence de densité de points n'affecte pas le calcul des métriques LiDAR de manière significative : sur les zones où les deux acquisitions se chevauchent (i.e. les zones frontalières des

départements 73 et 74 qui ont été couvertes par les deux campagnes), le calcul des mêmes métriques LiDAR avec l'une ou l'autre des sources de données montre une similitude dans les résultats. Par ailleurs concernant la différence de date, les deux acquisitions ont été réalisées à la même saison et ne sont espacées que de deux années : ce laps de temps est négligeable par rapport au temps caractéristique de variation des peuplements forestiers (de l'ordre de la décennie) et le PNR n'a pas subi d'aménagements sylvicoles majeurs dans cet intervalle. L'hypothèse d'une cohérence entre les relevés d'inventaire forestier réalisés en septembre 2018 et les campagnes d'acquisition LiDAR peut donc être retenue.

Des métriques LiDAR sont calculées pour chacune des 291 placettes de la BDD terrain. Le géoréférencement précis des placettes permet d'extraire les sous-nuages de points LiDAR correspondants dans les données brutes, soit tous les points contenus dans les emprises circulaires de rayon 15 m et centrées sur les coordonnées des centres des placettes. Les métriques calculées décrivent des caractéristiques des sous-nuages de points LiDAR correspondant à chaque placette et constituent les variables prédictives (variables X) des attributs de structure à prédire (variables Y). Au total 55 métriques différentes ont été calculées (cf. annexe 6.2) et peuvent être rangées dans les catégories suivantes :

- Métriques de caractérisation de la canopée : elles caractérisent la forme de l'environnement décrit par le nuage de points. Le taux de trouée (*gap fraction* en anglais) renseigne sur la pénétration de la lumière à travers la canopée et donc indirectement sur le taux de couvert. La rugosité de surface (*rumple index* en anglais), définie comme le rapport entre la surface de la canopée et cette surface projetée au sol, décrit la complexité de la structure de la canopée.
- Métriques de hauteur : elles caractérisent l'hétérogénéité verticale de la végétation. A partir de la distribution de hauteurs des points LiDAR plusieurs métriques standards sont calculées : hauteur maximale, hauteur moyenne, dispersion, asymétrie et aplatissement, percentiles de hauteur, ... Selon l'information recherchée tous les échos sont pris en compte, ou bien simplement les premiers retours, ou encore les derniers uniquement.
- Métriques d'intensité : elles caractérisent l'interaction entre le signal laser et l'environnement. Connaissant l'intensité rétrodiffusée et mesurée pour chaque écho LiDAR plusieurs métriques standards sont calculées selon le même principe que pour les métriques de hauteur : intensité totale, intensité moyenne, dispersion, asymétrie et aplatissement, intensités cumulées par percentiles de hauteur, ... Là encore tout ou partie des échos peuvent être pris en compte dans le calcul.

Chacune des 291 placettes terrain est donc décrite d'une part par des variables à prédire (attributs de structure, variables Y) et d'autre part par des variables prédictives (métriques LiDAR, variables X), permettant ainsi la calibration de modèles empiriques de prédiction.

4 DE LA CARTOGRAPHIE DE LA RESSOURCE FORESTIERE DU PNR DES BAUGES

4.1 Méthodologie

Les données LiDAR sont particulièrement appropriées pour l'estimation des attributs forestiers de structure. L'ONF a déjà produit de premières cartographies à partir de données LiDAR seules en utilisant une approche à l'arbre telle que décrite plus haut dans ce document. L'objectif final de cette mission professionnelle SILAT est d'évaluer si la complémentarité des informations optiques (SENTINEL-2) et structurales (LiDAR) permet d'améliorer la cartographie des attributs forestiers. Le format et la résolution des données radiométriques satellitaires ne permettent pas d'envisager une approche à l'individu et sont plus adaptées aux approches surfaciques.

Dans un premier temps des modèles à l'échelle de la placette ont donc été calibrés à partir des données LiDAR seules. Le but de cette première étape est d'explorer diverses méthodes empiriques, notamment par apprentissage, et de faire une évaluation comparative de leurs performances respectives. Cette étude complète l'approche à l'arbre déjà testée par l'ONF et apporte une nouvelle perspective à l'utilisation des données LiDAR dans le projet Protest. La mise en place de modèles empiriques débute par l'association des variables forestières de la vérité terrain avec les données de télédétection correspondantes pour créer une base d'apprentissage. Cette base est constituée des 291 objets « placette » décrits par leurs variables X/Y et permet de calibrer des modèles suivant diverses techniques. Dans le cadre de ce travail des modèles de régression construits par apprentissage automatique (*machine learning* en anglais) ont été évalués.

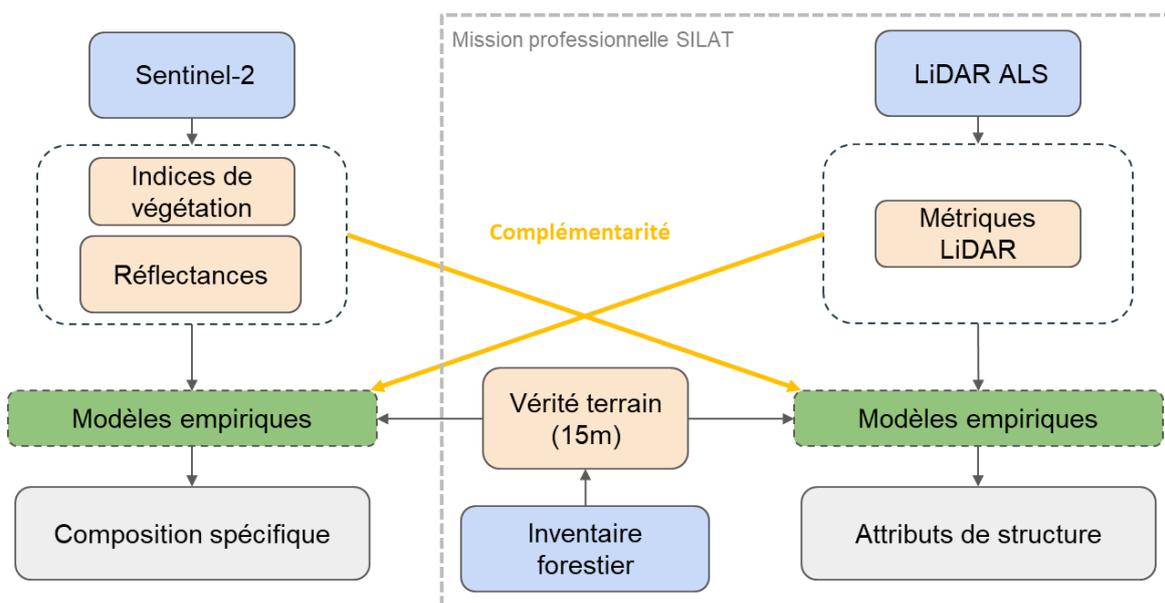


Figure 11 – Synthèse de la méthodologie

En parallèle un travail à la philosophie similaire est mené par un stagiaire de fin d'études de M2 sur la prédiction des variables d'état de la végétation (composition spécifique) à partir des données SENTINEL-2 seules. Ensuite, dans un second temps, le but est de faire converger ces deux travaux afin d'évaluer les perspectives offertes par la fusion de données et la création de modèles multi-sources : sont ainsi étudiés d'un côté l'apport des données radiométriques aux modèles d'estimation des attributs de structure (problématique de la mission professionnelle SILAT), et de l'autre celui des données LiDAR à la caractérisation des compositions spécifiques (problématique du stage de M2). Pour ce faire les modèles *deep learning* étudiés dans l'approche préalable LiDAR seul sont adaptés pour intégrer en entrée les données radiométriques. La pertinence et l'intérêt de l'ajout de ces données sont évalués en comparant les résultats obtenus avec les données LiDAR seules et ceux obtenus par fusion de données LiDAR et SENTINEL-2. Par ailleurs l'apport d'autres données exogènes comme les informations de topographie est également évalué. Enfin, une fois le meilleur modèle identifié, il est appliqué à toute la zone d'étude du PNR des Bauges pour produire les cartographies de la ressource forestière.

4.2 A partir de données LiDAR seules

Pour lier les attributs forestiers à cartographier avec les métriques LiDAR calculées à partir des nuages de points 3D, la création de modèles de régression par apprentissage automatique est la piste explorée dans ce projet. Les méthodes par apprentissage automatique consistent à entraîner des modèles de manière empirique à partir d'un jeu de données appelé « base d'apprentissage ». Le principe général consiste à rassembler un certain nombre d'exemples (i.e. la base d'apprentissage) pour déterminer des schémas sous-jacents (i.e. les modèles empiriques), puis à les utiliser pour effectuer des prévisions concernant de nouveaux exemples (i.e. la cartographie de toute la zone d'étude).

Dans le cadre de ce travail la base d'apprentissage est constituée de 291 objets placettes décrits par leurs descripteurs métriques LiDAR (variables X) et leurs attributs forestiers de structure (variables Y). L'approche par apprentissage automatique permet de créer des modèles de régression qui lient les variables Y aux variables X par entraînement à partir des données de la base d'apprentissage. Pour ce faire cette dernière est divisée selon trois jeux de données dont le rôle et la finalité sont bien définis :

- Le jeu de données « train » : constitué par environ 60% de la base d'apprentissage, soit 199 objets placettes, il sert à l'entraînement des modèles par apprentissage automatique. Le jeu de données « train » permet donc la calibration des paramètres internes des modèles empiriques lors du processus d'apprentissage.
- Le jeu de données « validation » : constitué par environ 20% de la base d'apprentissage, soit 59 objets placettes, il sert à fixer les valeurs d'hyperparamètres des modèles. Un hyperparamètre est un paramètre dont la valeur est utilisée pour contrôler le processus d'apprentissage. Il s'agit de paramètres généraux de définition des modèles qui ne peuvent

être appris à partir des données d'entraînement car ils sont définis en amont du processus d'apprentissage lui-même. Le jeu de données « validation » permet donc d'optimiser les valeurs d'hyperparamètres des modèles entraînés par le jeu de données « train ».

- Le jeu de données « test » : constitué par environ 20% de la base d'apprentissage, soit 58 objets placettes, il sert de référence pour étudier la pertinence des modèles créés grâce aux deux distributions décrites plus haut. Le jeu de données « test » permet donc d'évaluer les performances des modèles obtenus par apprentissage.

Cette division de la base d'apprentissage en trois jeux de données distincts couvre les trois étapes principales de calibration d'un modèle par apprentissage automatique. Dans la suite des travaux présentés ici plusieurs méthodes différentes sont comparées entre elles : pour disposer de la même base de comparaison et hiérarchiser ces méthodes entre elles il est nécessaire d'avoir recours à la même distribution train-validation-test. Celle-ci est donc fixée en amont et demeure invariante pour chaque modèle testé. Trois méthodes différentes d'apprentissage automatique ont été utilisées pour calibrer des modèles et évaluer leurs performances respectives : k-NN, Random Forest et réseaux de neurones (perceptron multi-couches, *multi-layer perceptron MLP* en anglais). L'application de ces méthodes nécessite de fournir en entrée des données normalisées pour de meilleurs résultats : en conséquence les variables X et Y des 291 objets de la base d'apprentissage ont été mises à l'échelle entre 0 et 1 à partir des valeurs minimum et maximum constatées sur toute la zone d'étude. La base d'apprentissage normalisée est ensuite divisée en trois partitions train-validation-test en respectant les proportions évoquées plus haut.

L'algorithme des k plus proches voisins (k-NN) est un algorithme d'apprentissage automatique qualifié de « paresseux » car il n'apprend rien pendant la phase d'entraînement. Pour prédire la valeur d'une nouvelle donnée d'entrée (i.e. un vecteur variables X), l'algorithme va rechercher ses k voisins les plus proches dans le jeu de données « train » (en utilisant la distance euclidienne) et appliquera une moyenne pondérée de la variable Y associée à ces k voisins par la distance euclidienne mesurée. Plus un voisin est proche de la donnée d'entrée que l'on souhaite prédire, plus sa contribution à la valeur finale prédite sera importante. Dans un premier temps la valeur optimale de k est déterminée à partir du jeu de données « validation ». Plusieurs valeurs de k comprises entre 1 et 10 sont successivement utilisées pour prédire les variables Y du jeu de données « validation » à partir du jeu de données « train » et la valeur de k qui minimise l'erreur entre la prédiction et la réalité est celle qui est finalement retenue. Pour cette étude $k_{opt} = 7$. La valeur optimale étant identifiée, les 7 voisins les plus proches des données du jeu « test » sont identifiés et les variables Y associées sont calculées par moyenne pondérée. Le taux d'erreur entre les valeurs prédites et la réalité de la base d'apprentissage permet d'évaluer les performances du modèle. Pour quantifier ces performances trois indicateurs ont

été retenus : l'erreur absolue moyenne (MAE), l'erreur quadratique moyenne (RMSE) et le coefficient de détermination linéaire de Pearson (R^2). Les résultats sont présentés en figure 12.

L'algorithme Random Forest (RF) est un algorithme d'apprentissage automatique basé sur l'assemblage d'arbres de décision. Plutôt que d'avoir un estimateur complexe capable de produire une prédiction à partir de toutes les données, le Random Forest utilise plusieurs estimateurs simples ayant chacun une vision partielle du problème général. L'ensemble de ces estimateurs est ensuite réuni pour obtenir une vision globale du problème. Chaque arbre est indépendant et dispose d'une vision parcellaire du problème par un double tirage aléatoire : un tirage aléatoire avec remplacement sur les observations (i.e. les objets placettes) et un tirage aléatoire sur les variables (i.e. les métriques LiDAR, variables X). Tous les arbres de décision sont finalement assemblés pour produire une prédiction calculée comme la moyenne de tous les arbres. Un hyperparamètre fondamental de l'algorithme est donc le nombre d'arbres de décision à construire, noté *n_{tree}*. Il influence directement sur les performances de l'algorithme et la qualité de la prédiction. Le nombre de variables tirées parmi toutes les variables X, noté *m_{try}*, est aussi un hyperparamètre fondamental. Par défaut ce nombre est égal à la racine carrée du nombre total de variables. De manière analogue à la démarche adoptée pour l'algorithme k-NN, les hyperparamètres optimaux sont déterminés à partir du jeu de données « validation ». Plusieurs valeurs d'hyperparamètres sont successivement utilisées pour entraîner un algorithme Random Forest à partir du jeu de données « train » et les valeurs qui minimisent l'erreur sur la prédiction des variables Y du jeu de données « validation » sont identifiées. Pour cette étude $n_{tree_{opt}} = 600$ et $m_{try_{opt}} = 7$. Un algorithme Random Forest défini par ces hyperparamètres optimaux est donc entraîné afin de prédire les variables Y du jeu de données « test ». Les performances du modèle sont évaluées de la même manière que précédemment (cf. résultats en figure 12).

Les réseaux de neurones artificiels sont des systèmes inspirés du fonctionnement biologique du cerveau. Le plus simple d'entre eux est le perceptron multi-couches (MLP) et c'est aussi celui qui est adapté au format de la base d'apprentissage : les données sont en effet présentées sous forme tabulaire, chaque ligne correspondant à un objet placette décrit par ses variables X et dont les variables Y sont connues par la vérité terrain. A l'image du fonctionnement biologique du cerveau le MLP est un ensemble de neurones organisés en couches. Pendant la phase d'entraînement, le signal d'entrée constitué par les variables X du jeu de données « train » se propage d'une couche à l'autre jusqu'à la sortie, en activant ou non au fur et à mesure des neurones. Le principe est de comparer la sortie obtenue par le MLP avec la sortie attendue et de mettre à jour les liaisons entre les neurones pour améliorer le résultat final. Ces liaisons se traduisent par des poids et un biais qui sont affectés au signal d'entrée de chaque neurone individuel. La sortie d'un neurone donné est quant à elle déterminée par une fonction d'activation appliquée à son signal d'entrée pondéré par les poids et le biais.

Concrètement cette activation représente le seuil à partir duquel le neurone va être « activé » et émettre un signal. La sortie de chaque neurone d'une couche est transmise à tous les neurones de la couche suivante du MLP. Du point de vue de l'architecture globale le MLP est ainsi organisé en trois parties successivement interconnectées entre elles : la couche d'entrée, la ou les couche(s) cachée(s) et la couche de sortie. Chaque couche contient un nombre donné de neurones. La couche d'entrée est un ensemble de neurones qui portent le signal d'entrée, les couches cachées permettent au réseau d'identifier les relations entre les variables et la couche de sortie produit la prédiction du réseau. Dans le cadre de ce projet l'architecture retenue est celle d'un MLP à trois couches dites « denses », c'est-à-dire entièrement connectées : une couche d'entrée, une couche cachée et une couche de sortie. Les attributs forestiers d'intérêt dans cette étude sont des variables continues et sont donc prédites par régression : en conséquence la couche de sortie du MLP est constituée d'un seul neurone qui produit en sortie une valeur de prédiction. Pour ce qui est des deux autres couches les nombres de neurones $N1$ et $N2$ sont deux hyperparamètres que l'on cherche à optimiser. La vitesse d'apprentissage lr_{rate} (*learning rate* en anglais) est également un autre hyperparamètre d'intérêt qui peut être optimisé. Pour cette étude $N1_{opt} = 256$, $N2_{opt} = 128$ et $lr_{rate}_{opt} = 0.001$.

Le processus d'apprentissage des paramètres du MLP (poids et biais des neurones) à partir des données « train » consiste à minimiser la perte entre la réponse obtenue en sortie du réseau et celle réellement attendue par un procédé appelé la « descente de gradient » : en partant de la sortie obtenue à une itération donnée du processus d'apprentissage (appelée epoch), une fonction de perte permet de quantifier l'erreur commise par le réseau. Les paramètres du réseau sont alors mis à jour à partir de cet écart, par rétropropagation de l'erreur en partant de la fin et en remontant le réseau vers l'amont. Dans cette étude le nombre d'epochs est fixé à 400. A chaque epoch les poids et biais des neurones sont mis à jour en fonction de la perte constatée en sortie. Cela signifie que le processus d'entraînement du MLP dure 400 itérations à l'issue desquelles la meilleure combinaison de paramètres du MLP est retenue pour l'établissement du modèle de prédiction. Les performances du modèle ainsi établi sont alors évaluées à partir du jeu de données « test ».

Pour chacune des trois méthodes décrites plus haut et pour les mêmes distributions train-validation-test de la base d'apprentissage, les performances des modèles de prédiction à partir de données LiDAR seules de la surface terrière, du volume de bois et du coefficient de Gini des DBH ont été évaluées. Les performances des modèles peuvent varier en fonction de la distribution choisie pour la base d'apprentissage. Par conséquent et afin de lisser les résultats, les modèles de prédiction ont été implémentés pour 30 distributions train-validation-test différentes tirées aléatoirement. Les métriques de performances sont ensuite moyennées sur ces 30 configurations. Les résultats sont présentés dans la figure 12.

Attribut	Méthode	MAE	RMSE	R ²
Surface terrière	k-NN	8,1 m ² /ha	10,9 m ² /ha	0,50
	RF	8,0 m ² /ha	10,8 m ² /ha	0,51
	MLP	7,9 m²/ha	10,4 m²/ha	0,54
Volume total	k-NN	91,9 m ³ /ha	127,3 m ³ /ha	0,52
	RF	87,7 m ³ /ha	124,5 m ³ /ha	0,54
	MLP	79,2 m³/ha	113,2 m³/ha	0,61
Gini des DBH	k-NN	0,039	0,053	0,21
	RF	0,036	0,048	0,29
	MLP	0,033	0,047	0,32

Figure 12 – Performances des modèles de prédiction à partir des données LiDAR seules

Les erreurs MAE et RMSE permettent de quantifier l'écart dans la prédiction des modèles, tandis que le coefficient R² est un indicateur qui mesure la qualité de la régression. L'association de ces trois métriques est donc pertinente pour décrire précisément les performances d'un modèle. La principale observation qui se dégage de ces résultats est que les attributs forestiers surface terrière et volume total sont bien décrits à partir de données LiDAR seules. En termes de performances le MLP est la méthode qui donne les meilleurs résultats pour les variables forestières étudiées dans ce projet. L'erreur quadratique moyenne sur la prédiction de la surface terrière est ainsi de 10,4 m²/ha et celle sur le volume est de 113,2 m³/ha. Une explication à ces performances améliorées pourrait tenir au fait que le MLP gère mieux un nombre élevé de variables en entrée que les autres méthodes. Aucune sélection préalable de métriques LiDAR par étude de leur corrélation n'a en effet été effectuée et les 55 métriques calculées sont utilisées comme variables X. Pour le coefficient de Gini des DBH, grandeur sans dimension comprise entre 0 et 1 en général, et entre 0 et 0,5 dans cette étude, l'erreur quadratique moyenne est de 0,047. Les valeurs de R² obtenues pour la prédiction de cet attribut sont nettement plus basses et témoignent donc d'un modèle de moins bonne qualité. Cela peut s'expliquer par le fait que les descripteurs LiDAR (variables X) utilisés pour la construction des modèles caractérisent en majorité l'hétérogénéité verticale de la forêt, tandis que le Gini des DBH est un attribut qui représente plutôt l'hétérogénéité horizontale d'un peuplement forestier. D'où un constat de performances inférieures par rapport à d'autres attributs comme la surface terrière et le volume total qui sont bien décrits par les métriques LiDAR calculées.

4.3 Apport de la fusion de données multi-sources

Les données LiDAR seules ont donc démontré leur pertinence pour la prédiction des attributs forestiers d'intérêt dans cette étude. La deuxième étape de ce travail telle que décrite dans la méthodologie consiste à enrichir les variables X des modèles avec de nouvelles données issues de

sources multiples et d'évaluer l'apport d'une approche de fusion. Compte tenu des résultats discutés dans la section précédente le choix est fait d'adopter une approche de fusion multi-sources à partir d'une architecture réseau de neurones car c'est celle qui a donné les meilleurs résultats.

Les données SENTINEL-2 ont fait l'objet d'un traitement complet dans le cadre d'un travail de stage de M2 mené en parallèle de cette mission professionnelle. Le téléchargement des données sur la zone d'étude et les prétraitements nécessaires ont ainsi déjà été effectués et ne sont pas évoqués dans le détail ici. De par leur format différent de celui des données LiDAR, elles font néanmoins l'objet d'une approche différente pour leur intégration dans les modèles de prédiction. En effet les données SENTINEL-2 se présentent sous la forme d'images acquises dans plusieurs bandes et à des dates d'acquisition distinctes. 10 bandes spectrales ont été utilisées parmi toutes celles disponibles : trois bandes dans le spectre visible, cinq dans le proche infrarouge et deux dans l'infrarouge court. La résolution spatiale des images varie en fonction de la bande spectrale selon trois valeurs : 10 mètres, 20 mètres et 60 mètres. Pour cette étude toutes les bandes ont été reprojctées à 10 mètres de résolution. En ce qui concerne les séries temporelles cinq acquisitions entre 2017 et 2018 ont été jugées pertinentes pour l'analyse (taux d'enneigement nul sur la zone du PNR des Bauges et absence de neige). Ainsi les données SENTINEL-2 utilisées pour l'établissement de nouveaux modèles de prédiction se présentent sous la forme d'une pile de 50 rasters géoréférencés à une résolution spatiale de 10 mètres (10 bandes \times 5 dates = 50 rasters).

Ce formalisme différent des données implique une hypothèse à prendre sur la définition d'un objet placette du point de vue des données SENTINEL-2. En effet, comme précisé plusieurs fois précédemment, les placettes d'inventaire sont formées par des cercles de rayon 15 mètres. Compte tenu de la résolution spatiale des données SENTINEL-2 (10 mètres), le choix a été fait d'extraire l'information spectrale sur des fenêtres de dimension 3 pixels sur 3 pixels, soit 30 mètres de côté qui correspondent au diamètre des placettes terrain. Le pixel central d'une fenêtre donnée est positionné sur le centre géographique de la placette. Ainsi du point de vue de SENTINEL-2 un objet placette est décrit par l'information spectrale d'une fenêtre carrée de 9 pixels et sur 50 bandes. Pour construire un modèle de prédiction à partir de données de ce type une architecture réseau de neurones convolutif (CNN) est particulièrement adaptée. La structure détaillée de ce type de réseau n'est pas présentée ici : le travail de stage de M2 mené en parallèle a mis l'accent sur cette approche et la même méthodologie a pu être mise en œuvre directement dans cette approche de fusion.

En plus des données radiométriques SENTINEL-2 des informations de topographie sur la zone d'étude ont pu être dérivées à partir des mesures LiDAR. Un modèle numérique de terrain (MNT) d'une résolution de 1 mètre a ainsi été produit par les prestataires responsables des campagnes d'acquisition LiDAR. Ce MNT a ensuite été reprojcté à une résolution de 10 mètres pour être

conforme à la résolution spatiale des données optiques. La pente, l'orientation et l'élévation ont été calculées à cette résolution spatiale sur la zone du PNR des Bauges, ce qui porte le nombre de rasters géoréférencés sur la zone à 53 (50 bandes spectrales, 3 bandes de topographie). Par ailleurs cette même information de topographie a aussi été déterminée à l'échelle de la placette : les moyennes de la pente, de l'orientation et de l'élévation sur les fenêtres de taille 9 pixels associées aux placettes ont ainsi été calculées pour produire une information unique de ces trois paramètres de topographie au niveau d'un objet placette.

L'enjeu est donc de développer une nouvelle approche de prédiction basée sur quatre sources de données rangées selon deux catégories :

- Données au niveau placette : métriques LiDAR et topographie moyennée (pente, orientation et élévation moyennées)
- Données au niveau pixel : bandes spectrales SENTINEL-2 (50 bandes) et topographie (3 bandes)

Ces deux catégories de données vont définir la structure du modèle de prédiction mis en place. L'architecture de type MLP est adaptée pour le traitement des données au niveau placette (données tabulaires) tandis que l'architecture CNN est celle retenue pour les données au niveau pixel (données rasters). Le modèle de prédiction par fusion de données multi-sources se présente donc sous la forme de deux branches distinctes (MLP d'une part, CNN d'autre part) dont les sorties intermédiaires sont combinées les unes avec les autres. Il est important de noter qu'à cette étape aucune prédiction n'est produite. Le résultat de cette combinaison est ensuite envoyé dans une branche finale qui synthétise la fusion et produit la prédiction. Afin de couvrir toutes les configurations possibles et de comprendre l'apport de chaque type de données, toutes les combinaisons possibles de données fournies en entrée du modèle ont été testées. Les distributions des placettes train-validation-test restent les mêmes que celles utilisées dans la section précédente pour pouvoir comparer les performances à conditions égales. Les résultats sont présentés dans la figure 13. En raison de performances très basses et de résultats non-interprétables pour la prédiction du coefficient de Gini des DBH, la suite de l'étude se concentre exclusivement sur les attributs surface terrière et volume total de bois.

Attribut	Données	MAE	RMSE	R ²
Surface terrière	LiDAR	7,9 m ² /ha	10,4 m ² /ha	0,54
	LiDAR + S2	7,4 m ² /ha	10,0 m ² /ha	0,59
	(LiDAR, topo)	7,0 m ² /ha	9,3 m ² /ha	0,65
	(LiDAR, topo) + S2	7,1 m ² /ha	9,6 m ² /ha	0,61
	LiDAR + (S2, topo)	6,3 m ² /ha	8,5 m ² /ha	0,71
	(LiDAR, topo) + (S2, topo)	6,0 m²/ha	8,4 m²/ha	0,72
Volume total	LiDAR	79,2 m ³ /ha	113,2 m ³ /ha	0,61
	LiDAR + S2	79,2 m ³ /ha	114,6 m ³ /ha	0,60
	(LiDAR, topo)	77,8 m ³ /ha	101,8 m ³ /ha	0,70
	(LiDAR, topo) + S2	80,6 m ³ /ha	104,7 m ³ /ha	0,68
	LiDAR + (S2, topo)	67,9 m ³ /ha	99,0 m ³ /ha	0,71
	(LiDAR, topo) + (S2, topo)	62,2 m³/ha	93,4 m³/ha	0,74

Figure 13 - Performances du modèle de prédiction par fusion en fonction des données d'entrée

L'ajout de nouvelles sources de données engendre une amélioration significative des résultats produits par le modèle de prédiction par rapport à l'utilisation des données LiDAR seules. Comme expliqué précédemment les données SENTINEL-2 seules ne sont pas adaptées pour la prédiction d'attributs forestiers de structure. Un modèle construit à partir de données SENTINEL-2 seules a ainsi été construit et n'a pas donné de résultats concluants (R² compris entre 0,2 et 0,3). L'association des données SENTINEL-2 avec les données LiDAR montre une amélioration limitée par rapport à l'approche LiDAR seul : une légère amélioration est constatée pour l'estimation de la surface terrière tandis que les résultats sont équivalents pour la prédiction du volume total. En revanche les données topographiques apportent une véritable contribution à la qualité des modèles. Les données topographiques calculées au niveau pixel (c.-à-d. à une résolution spatiale de 10 mètres) sont celles dont l'impact est le plus significatif : l'association des données LiDAR avec les données SENTINEL-2 incluant la topographie entraîne une diminution des erreurs constatées et une amélioration des modèles (le R² passe de 0,54 à 0,71 pour la surface terrière et de 0,61 à 0,71 pour le volume total). De la même manière, mais dans une moindre mesure, les données de topographie moyenne calculées à l'échelle de la placette apportent aussi une amélioration par rapport au LiDAR seul. En résumé, si l'impact des données SENTINEL-2 reste difficile à constater de manière significative, celui des données topographiques est très clair. L'approche multi-sources qui combine toutes les données disponibles apparaît alors comme la meilleure configuration possible : celle-ci permet d'atteindre des R² de l'ordre de 0,7 pour la surface terrière et le volume total. La duplication des mêmes données de topographie mais calculées à différentes échelles (placette et pixel) contribue positivement aux résultats produits. Des représentations graphiques sous la forme prédit vs. mesuré sont données en annexes 6.4 pour

apprécier visuellement la qualité de la prédiction du modèle multi-sources avec toutes les données disponibles.

L'apport significatif des informations de topographie pourrait tenir au fait que cette dernière impacte directement les mesures de télédétection, qu'elles soient aéroportées ou satellitaires. Les nuages de points ALS utilisés dans le cadre de cette étude ont été normalisés en hauteur. Cela signifie que l'élévation du sol (i.e. le relief) a été soustraite à l'altitude absolue mesurée, et ce afin de disposer de données cohérentes pour le calcul des différentes métriques ALS. Ce processus induit une distorsion des nuages de points LiDAR causée par les différences d'élévation, et donc une variation dans le calcul des métriques associées. La prise en compte des informations de topographie dans les modèles pourrait donc apporter une correction à cet effet lié à l'acquisition des données. Ce même constat peut également être posé pour les données optiques SENTINEL-2. La réponse spectrale de l'environnement observé est en effet sensible aux effets de relief. Une exposition différente à la lumière du soleil (par exemple pour deux versants d'une même montagne) impacte directement la réflectance mesurée par les capteurs. Même si ces biais sont théoriquement pris en compte lors de l'élaboration des produits SENTINEL-2, l'ajout des informations de topographie à l'échelle des bandes spectrales SENTINEL-2 dans les modèles de prédiction pourrait agir comme un élément de correction supplémentaire et permettre un gain de performances.

4.4 Cartographies de la ressource forestière

Le modèle retenu pour l'établissement des cartographies de la ressource forestière du PNR des Bauges est donc une architecture de type réseau de neurones divisée en deux branches qui permettent d'adopter une approche de fusion de données multi-sources : à partir de données à l'échelle de la placette d'une part (LiDAR et topographie moyenne) et de données au niveau pixel de 10 mètres d'autre part (SENTINEL-2 et topographie). Cette différence d'échelles impose une contrainte pour la résolution spatiale des cartographies réalisées. Compte tenu de la dimension d'une placette (cercle de 30 mètres de diamètre), le choix est fait de travailler à une résolution spatiale de 30 mètres. L'ensemble de la zone d'étude est donc découpé selon une grille régulière de résolution 30 mètres et le modèle est appliqué sur chaque pixel de cette grille. Les cartes ainsi obtenues de la ressource forestière du PNR des Bauges sont données ci-après. L'attribut coefficient de Gini des DBH n'a pas été cartographié en raison des résultats peu satisfaisants obtenus lors de l'évaluation des modèles de prédiction.

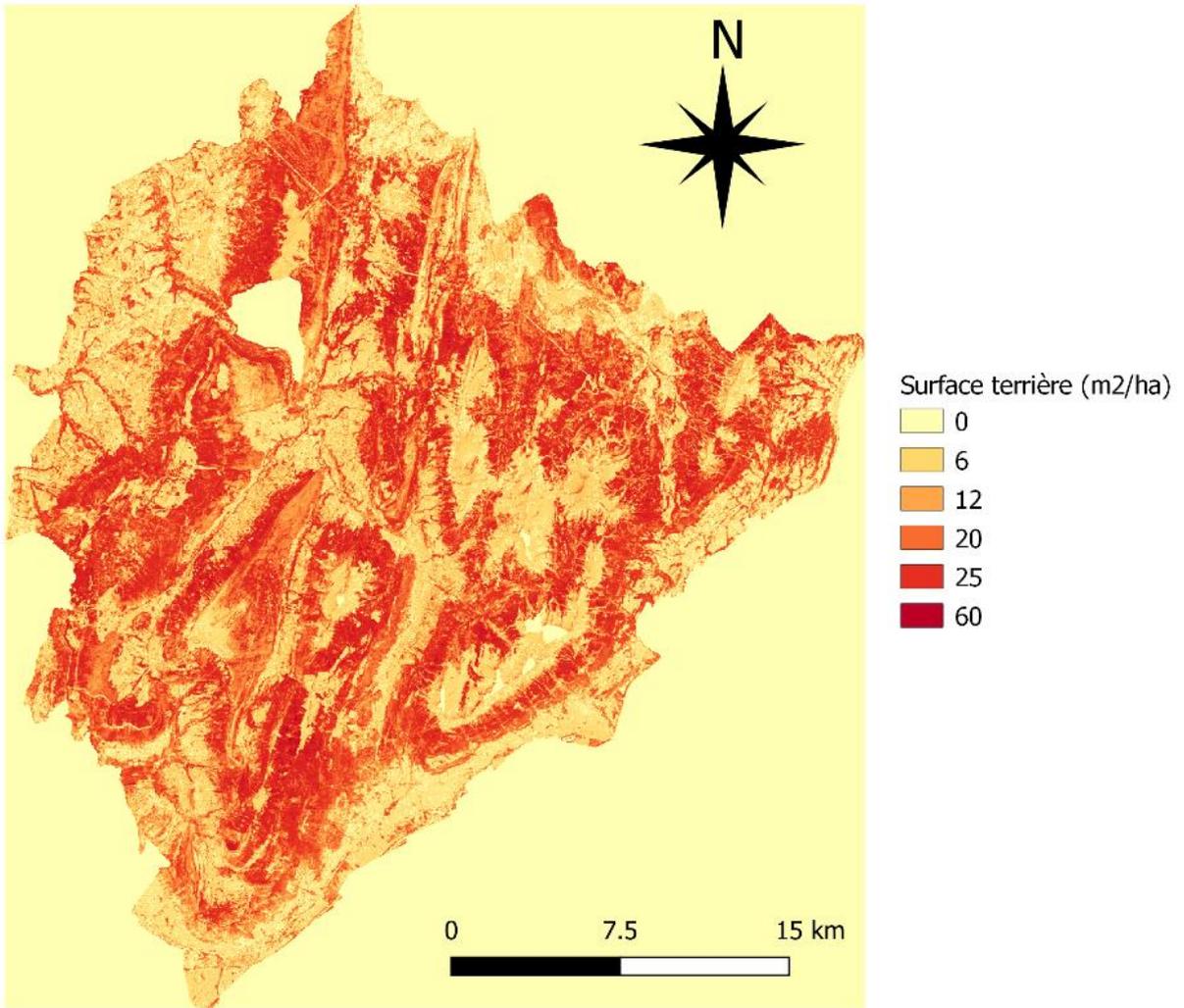


Figure 14 - Cartographie de la surface terrière du PNR des Bauges (résolution 30 m)

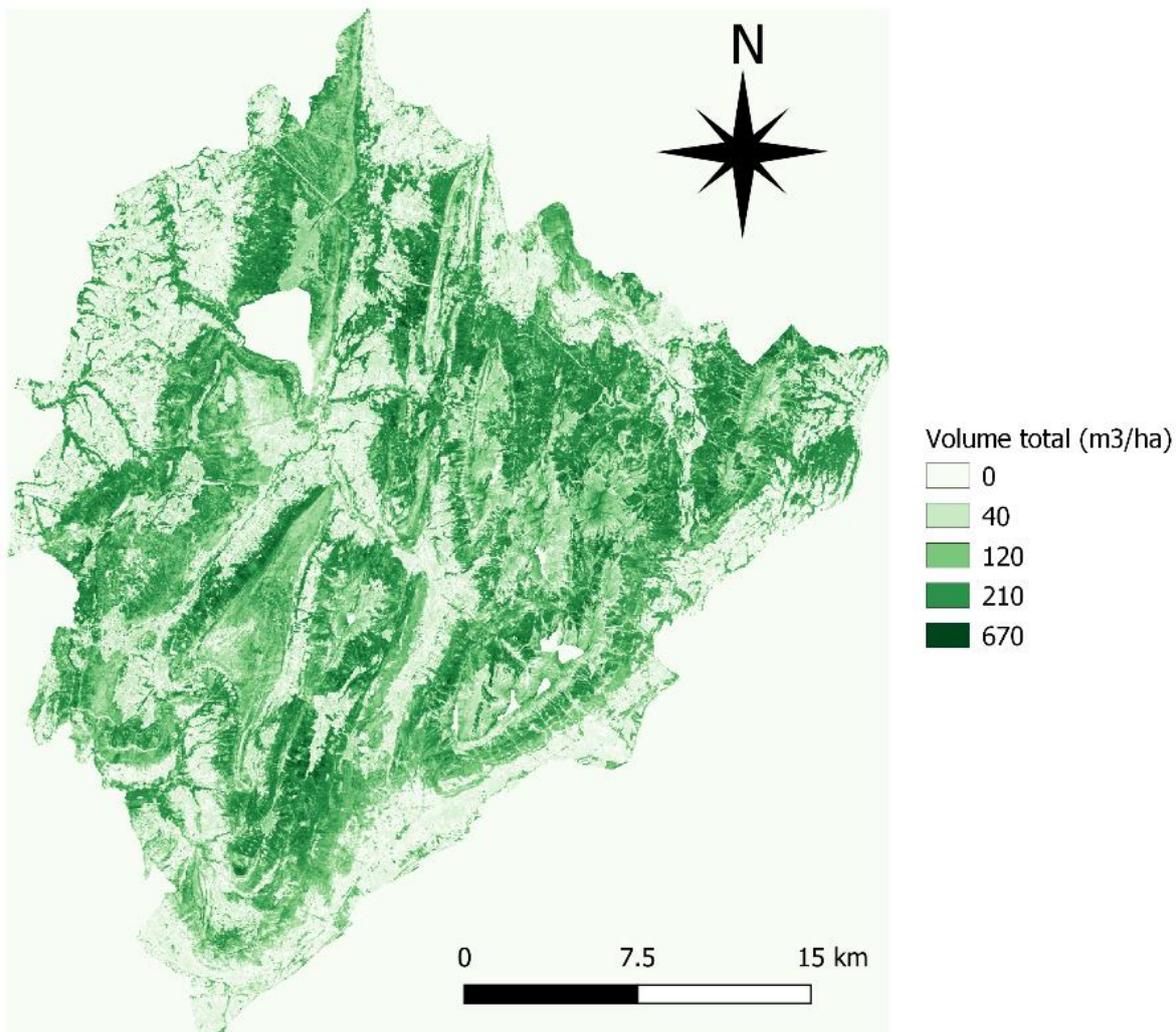


Figure 15 - Cartographie du volume total de bois du PNR des Bauges (résolution 30 m)

Ces cartographies indiquent un lien apparent entre les deux attributs surface terrière et volume total, ce qui est effectivement le cas compte tenu du fait qu'ils sont tous les deux liés à une mesure quantitative de la biomasse de bois. La comparaison de ces cartes avec d'autres résultats obtenus dans le cadre du projet Protest par une méthode alternative (approche à l'arbre) montre des tendances analogues et permet de valider la pertinence relative des résultats obtenus.

5 CONCLUSION

Les travaux menés dans le cadre de ce projet ont permis de démontrer l'intérêt de l'utilisation de différentes sources de données de télédétection pour l'amélioration de la prédiction de la surface terrière et du volume total de bois. Si le LiDAR seul reste une technologie particulièrement adaptée pour décrire la structure des écosystèmes forestiers, son association avec d'autres sources de données externes a permis d'améliorer significativement les performances des modèles de prédiction. En particulier les données topographiques ont contribué à la diminution des erreurs constatées et à l'amélioration de la qualité des modèles. Leur pertinence est d'autant plus importante que la zone d'étude du PNR des Bauges se caractérise par une forte hétérogénéité dans sa topographie. L'impact positif de l'ajout des données SENTINEL-2 a aussi été constaté dans une moindre mesure : bien que les données radiométriques soient plus adaptées à la description de l'état de la végétation, elles ont néanmoins démontré une certaine pertinence dans les approches empiriques par apprentissage automatique présentées ici.

La fusion de données multi-sources dans une architecture de type réseau de neurones a donné l'opportunité de faire converger des informations de natures différentes et à des échelles multiples. Le modèle de cartographie élaboré au cours de cette mission professionnelle se caractérise ainsi par deux branches distinctes dédiées chacune au traitement d'un type particulier de données : d'une part une branche traitant les informations à l'échelle de la placette (métriques LiDAR et topographie moyennée), et d'autre part une branche traitant les informations à l'échelle du pixel (données optiques SENTINEL-2 et topographie). Les approches alternatives présentées dans ce rapport (k-NN et RF) sont par définition limitées en ce qui concerne la prise en compte de plusieurs sources de données hétérogènes. La quantité de données utilisées importe donc autant que la manière dont elles sont fusionnées. L'approche par apprentissage profond est celle qui a permis de concilier ces deux aspects fondamentaux dans l'élaboration d'un modèle de prédiction multi-sources.

L'application des modèles ainsi établis à toute la zone d'étude a permis de générer deux cartes de la ressource forestière du PNR des Bauges à travers la représentation de deux attributs forestiers de structure à une échelle de 30 mètres : la surface terrière et le volume de bois. Ce type de produits cartographiques s'inscrit dans le cadre du lot « Cartographie » du projet Protest et constitue une contribution aux objectifs d'estimation de la ressource forestière du PNR. La pertinence de l'approche empirique par apprentissage qui a été adoptée pour la réalisation de ces cartographies est directement liée au fait de disposer de mesures terrain justes et suffisantes. Le nombre de placettes d'inventaire ainsi que la complétude des mesures dendrométriques effectuées sont donc deux éléments fondamentaux qui ont un impact significatif sur les performances des modèles établis, au même titre que les données de télédétection utilisées ou l'approche méthodologique retenue.

6 ANNEXES

6.1 Base de données terrain

Les travaux présentés dans la section 2 ont permis l'élaboration d'une base de données (BDD) terrain exhaustive à l'échelle de la placette. Les champs contenus dans cette BDD sont décrits dans le dictionnaire ci-dessous :

Nom	Type	Description
Id_plac	String	Identifiant de la placette
X	Float	Coordonnée X du centre de la placette
Y	Float	Coordonnée Y du centre de la placette
dep	Integer	Département correspondant à l'acquisition LiDAR
BDForet	String	Essence(s) dominante(s) issues de la BDForet
stratum	String	Typologie d'espèce(s) de la placette à partir du champ essence de la BDForet : Feuillus, Conifères ou Mixte
Brin_taill	Integer	Nombre de brins de taillis dans la placette de 10m : diamètre entre 7,5 et 17,5 cm
pch.cl15.F	Integer	Nombre de perches viables feuillues dans la placette de 10m : diamètre entre 7,5 et 17,5 cm
pch.cl15.R	Integer	Nombre de perches viables résineuses dans la placette de 10m : diamètre entre 7,5 et 17,5 cm
perch.nv	Integer	Nombre de perches non viables dans la placette de 10m : diamètre entre 7,5 et 17,5 cm
nb_arbre_175	Integer	Nombre d'arbres de diamètre supérieur à 17,5 cm dans la placette de 15m
nb_arbre_175_F	Integer	Nombre d'arbres feuillus de diamètre supérieur à 17,5 cm dans la placette de 15m
nb_arbre_175_R	Integer	Nombre d'arbres résineux de diamètre supérieur à 17,5 cm dans la placette de 15m
Brin_taill.norm	Integer	Nombre de brins de taillis normalisé dans la placette de 15m
pch.cl15.F.norm	Integer	Nombre de perches viables feuillues normalisé dans la placette de 15m
pch.cl15.R.norm	Integer	Nombre de perches viables résineuses normalisé dans la placette de 15m
perch.nv.norm	Integer	Nombre de perches non viables normalisé dans la placette de 15m
perch.nv.F.norm	Integer	Nombre estimé de perches non viables feuillues normalisé dans la placette de 15m
perch.nv.R.norm	Integer	Nombre estimé de perches non viables résineuses normalisé dans la placette de 15m
nb_arbre	Integer	Nombre total d'arbres normalisé dans la placette de 15m
comp_F_nb	Float	Pourcentage en nombre d'individus d'arbres feuillus dans la placette de 15m
comp_R_nb	Float	Pourcentage en nombre d'individus d'arbres résineux dans la placette de 15m
comp_nb	String	Typologie d'espèce(s) de la placette de 15m en nombre

		d'individus : Feuillus, Résineux ou Mixte
G175	Float	Surface terrière totale placette 15m des arbres de diamètre supérieur à 17,5 cm
G175R	Float	Surface terrière totale placette 15m des arbres résineux de diamètre supérieur à 17,5 cm
G175F	Float	Surface terrière totale placette 15m des arbres feuillus de diamètre supérieur à 17,5 cm
Gbt	Float	Surface terrière totale placette 15m des brins de taillis
GpvR	Float	Surface terrière totale placette 15m des perches viables résineuses
GpvF	Float	Surface terrière totale placette 15m des perches viables feuillues
GpvnF	Float	Surface terrière totale placette 15m des perches non viables feuillues
GpvnR	Float	Surface terrière totale placette 15m des perches non viables résineuses
G75	Float	Surface terrière totale placette 15m
comp_F_G	Float	Pourcentage en surface terrière d'arbres feuillus dans la placette de 15m
comp_R_G	Float	Pourcentage en surface terrière d'arbres résineux dans la placette de 15m
comp_G	String	Typologie d'espèce(s) de la placette de 15m en surface terrière : Feuillus, Résineux ou Mixte
gini_diam_median	Float	Coefficient Gini de la distribution des diamètres des arbres de la placette de 15m en prenant une valeur de diamètre fixe pour les perches/brins (médiane IFN i.e. 11,1 cm)
lorenz_diam_median	Float	Coefficient d'asymétrie de Lorenz de la distribution des diamètres des arbres de la placette de 15 m en prenant une valeur de diamètre fixe pour les perches/brins (médiane IFN i.e. 11,1 cm)
gini_diam_175	Float	Coefficient Gini de la distribution des diamètres des arbres de diamètre supérieur à 17,5 cm
lorenz_diam_175	Float	Coefficient d'asymétrie de Lorenz de la distribution des diamètres des arbres de diamètre supérieur à 17,5 cm
ratio	Float	Ratio du nombre de perches/brins sur le nombre total d'arbres dans la placette
petit_bois	Float	Pourcentage en nombre d'individus de petit bois dans la placette ($D < 22.5$ cm)
moyen_bois	Float	Pourcentage en nombre d'individus de moyen bois dans la placette ($22.5 \text{ cm} < D < 47.5$ cm)
gros_bois	Float	Pourcentage en nombre d'individus de gros bois dans la placette ($47.5 \text{ cm} < D < 67.5$ cm)
tres_gros_bois	Float	Pourcentage en nombre d'individus de très gros bois dans la placette ($D > 67.5$ cm)
gini_diam_rand	Float	Coefficient Gini de la distribution des diamètres des arbres de la placette de 15m en prenant une valeur aléatoire de diamètre pour chaque perche/brin (tirage uniforme dans l'intervalle [7,5 ; 17,5])

lorenz_diam_rand	Float	Coefficient d'asymétrie de Lorenz de la distribution des diamètres des arbres de la placette de 15m en prenant une valeur aléatoire de diamètre pour chaque perche/brin (tirage uniforme dans l'intervalle [7,5 ; 17,5])
gini_diam	Float	Coefficient Gini de la distribution des diamètres des arbres de la placette de 15m (sélection de la valeur gini_diam_median ou gini_diam_rand en fonction de la valeur du champ ratio)
lorenz_diam	Float	Coefficient d'asymétrie de Lorenz de la distribution des diamètres des arbres de la placette de 15m (sélection de la valeur lorenz_diam_median ou lorenz_diam_rand en fonction de la valeur du champ ratio)
density	Float	Nombre d'arbres par hectare dans la placette de 15m
htot_moy	Float	Hauteur moyenne dans la placette de 15m
htot_dom	Float	Hauteur dominante dans la placette de 15m (hauteur moyenne des 7 arbres les plus grands)
htot_Lorey	Float	Hauteur de Lorey dans la placette de 15m (hauteur moyenne pondérée par la surface terrière)
volume_total	Float	Volume total de la placette de 15m (somme des volumes totaux des arbres de la placette)
volume_total_175	Float	Volume total des arbres de diamètre supérieur à 17,5 cm
volume_tige	Float	Volume de tige de la placette (somme des volumes de tige des arbres de la placette)
volume_tige_175	Float	Volume de tige des arbres de diamètre supérieur à 17,5 cm

6.2 Métriques LiDAR

Au total 55 métriques LiDAR ont été calculées à l'échelle de la placette. La liste exhaustive de ces métriques est donnée ci-dessous :

Nom	Type	Description
gap_F	Float	Taux de trouée (gap fraction) de la placette calculée selon la méthode du nuage de points
rumple_index	Float	Rumple index de la placette
zmax	Float	Maximum des hauteurs des 1ers retours LiDAR
zmean	Float	Moyenne des hauteurs des 1ers retours LiDAR
zsd	Float	Dispersion (écart-type) des hauteurs des 1ers retours LiDAR
zq*	Float	Percentiles de hauteur (*%, * entier compris entre 5 et 95 par pas de 5)
skewness_first	Float	Skewness des hauteurs des 1ers retours LiDAR
kurtosis_first	Float	Kurtosis des hauteurs des 1ers retours LiDAR
skewness_last	Float	Skewness des hauteurs des derniers retours LiDAR
kurtosis_last	Float	Kurtosis des hauteurs des derniers retours LiDAR
imean	Float	Intensité moyenne normalisée des premiers retours LiDAR
zentropy	Float	Entropie de la distribution de hauteur des premiers retours LiDAR
pzabovemean	Float	Pourcentage de retours au-dessus de la hauteur moyenne
itot	Float	Intensité totale normalisée des premiers retours LiDAR

imax	Float	Intensité maximale normalisée des premiers retours LiDAR
isd	Float	Dispersion (écart-type) des intensités des premiers retours LiDAR
iskew	Float	Skewness des intensités des premiers retours LiDAR
ikurt	Float	Kurtosis des intensités des premiers retours LiDAR
zpcum*	Float	Pourcentage cumulé de retours LiDAR dans la couche * (couches de 1/10ème de zmax)
ipcumzq**	Float	Pourcentage d'intensité totale retournée sous le ** percentile de hauteur (** entier compris entre 10 et 90 par pas de 20)
p*th	Float	Pourcentage de *-ème retours (* entier compris entre 1 et 5)

6.3 Architecture du modèle à 2 branches

Le modèle de fusion multi-sources utilisé pour produire la cartographie de la ressource forestière du PNR des Bauges est constitué de deux branches dédiées chacune au traitement d'une échelle spécifique des données (placette d'une part, pixel d'autre part). L'architecture globale est décrite dans la figure 16 :

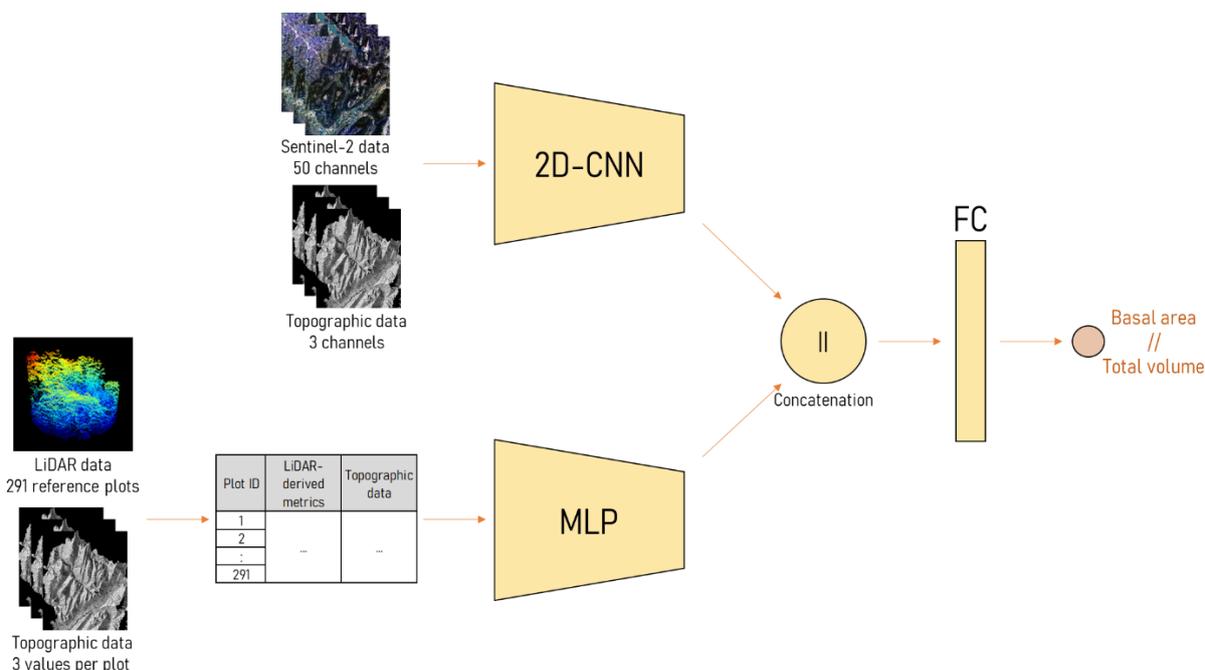


Figure 16 - Architecture du modèle à deux branches

6.4 Performances du modèle à 2 branches

Un jeu de données « test » est constitué de 50 placettes et les performances des modèles ont été évaluées à partir de 30 distributions différentes. Par conséquent, 1500 prédictions peuvent être représentées graphiquement en comparaison avec les valeurs réellement mesurées et donner une appréciation visuelle des modèles de prédiction pour les deux variables forestières :

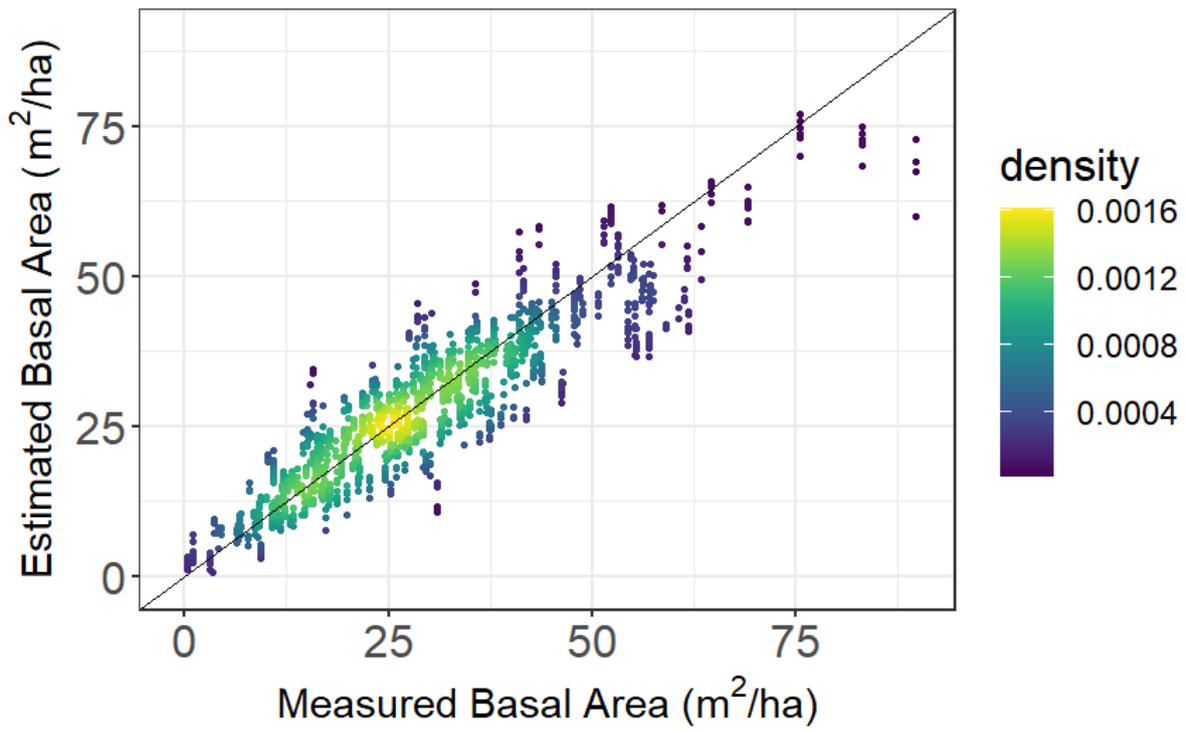


Figure 17 - Surface terrière mesurée vs. estimée à partir du modèle multi-sources

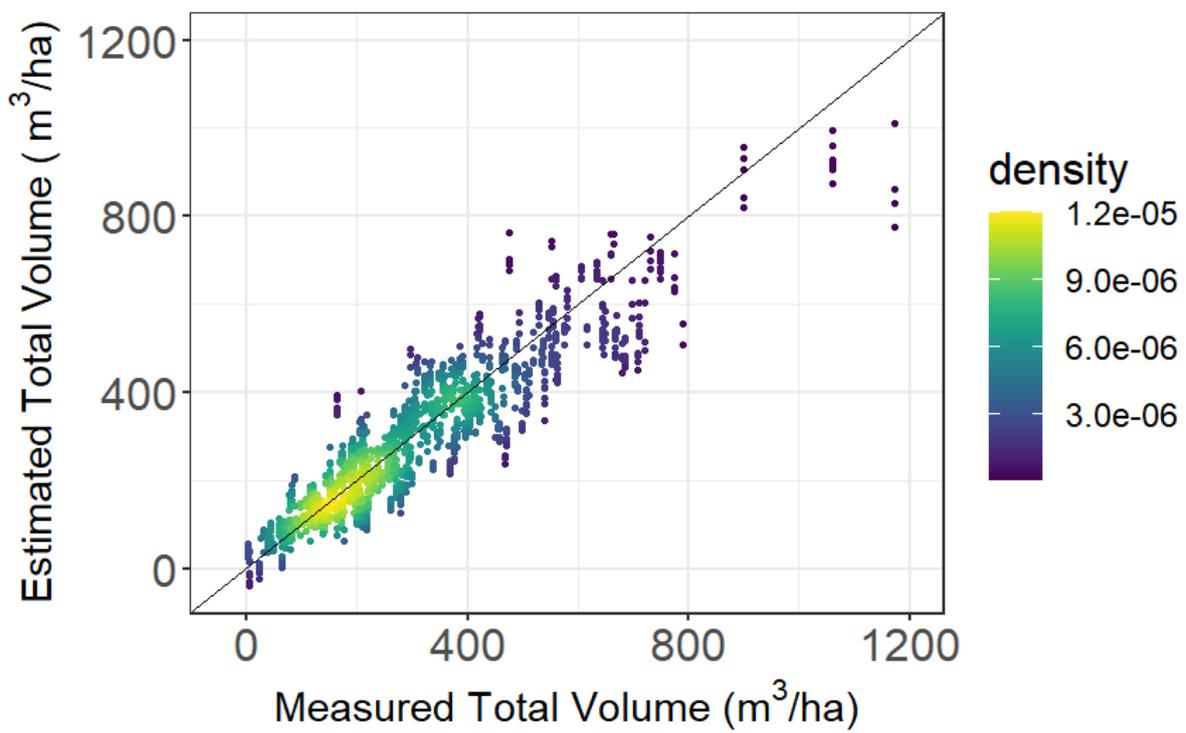


Figure 18 - Volume total mesuré vs. estimé à partir du modèle multi-sources

7 BIBLIOGRAPHIE

- Beer, C., M. Reichstein, E. Tomelleri, P. Ciais, M. Jung, N. Carvalhais, C. Rodenbeck, et al. 2010. « Terrestrial Gross Carbon Dioxide Uptake: Global Distribution and Covariation with Climate ». *Science* 329 [5993]: 834-38. <https://doi.org/10.1126/science.1184984>.
- Benedetti, Paola, Dino Ienco, Raffaele Gaetano, Kenji Ose, Ruggero G. Pensa, et Stéphane Dupuy. 2018. « M^3 Fusion: A Deep Learning Architecture for Multiscale Multimodal Multitemporal Satellite Data Fusion ». *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11 [12]: 4939-49. <https://doi.org/10.1109/JSTARS.2018.2876357>.
- Bouvier, Marc, Sylvie Durrieu, Richard A. Fournier, et Jean-Pierre Renaud. 2015. « Generalizing Predictive Models of Forest Inventory Attributes Using an Area-Based Approach with Airborne LiDAR Data ». *Remote Sensing of Environment* 156 [janvier]: 322-34. <https://doi.org/10.1016/j.rse.2014.10.004>.
- Boyd, D. S., et F. M. Danson. 2005. « Satellite Remote Sensing of Forest Resources: Three Decades of Research Development ». *Progress in Physical Geography: Earth and Environment* 29 [1]: 1-26. <https://doi.org/10.1191/0309133305pp432ra>.
- Cabello, Javier, Néstor Fernández, Domingo Alcaraz-Segura, Cecilio Oyonarte, Gervasio Piñeiro, Alice Altesor, Miguel Delibes, et José M. Paruelo. 2012. « The Ecosystem Functioning Dimension in Conservation: Insights from Remote Sensing ». *Biodiversity and Conservation* 21 [13]: 3287-3305. <https://doi.org/10.1007/s10531-012-0370-7>.
- Chave, J., C. Andalo, S. Brown, M. A. Cairns, J. Q. Chambers, D. Eamus, H. Fölster, et al. 2005. « Tree Allometry and Improved Estimation of Carbon Stocks and Balance in Tropical Forests ». *Oecologia* 145 [1]: 87-99. <https://doi.org/10.1007/s00442-005-0100-x>.
- DeFries, Ruth, Andrew Hansen, B. L. Turner, Robin Reid, et Jianguo Liu. 2007. « LAND USE CHANGE AROUND PROTECTED AREAS: MANAGEMENT TO BALANCE HUMAN NEEDS AND ECOLOGICAL FUNCTION ». *Ecological Applications* 17 [4]: 1031-38. <https://doi.org/10.1890/05-1111>.
- Franklin, Jerry F., Kermit Jr. Cromack, William Denison, Arthur McKee, Chris Maser, James Sedell, Fred Swanson, et Glen Juday. 1981. « Ecological Characteristics of Old-Growth Douglas-Fir Forests. » PNW-GTR-118. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Forest and Range Experiment Station. <https://doi.org/10.2737/PNW-GTR-118>.
- García-Gutiérrez, Jorge, Eduardo González-Ferreiro, Daniel Mateos-García, et José C. Riquelme-Santos. 2016. « A Preliminary Study of the Suitability of Deep Learning to Improve LiDAR-Derived Biomass Estimation ». In *Hybrid Artificial Intelligent Systems*, édité par Francisco Martínez-Álvarez, Alicia Troncoso, Héctor Quintián, et Emilio Corchado, 9648:588-96. Lecture Notes in Computer Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-32034-2_49.
- Grabska, Ewa, Patrick Hostert, Dirk Pflugmacher, et Katarzyna Ostapowicz. 2019. « Forest Stand Species Mapping Using the Sentinel-2 Time Series ». *Remote Sensing* 11 [10]: 1197. <https://doi.org/10.3390/rs11101197>.
- Ienco, Dino, Roberto Interdonato, Raffaele Gaetano, et Dinh Ho Tong Minh. 2019. « Combining Sentinel-1 and Sentinel-2 Satellite Image Time Series for Land Cover Mapping via a Multi-Source Deep Learning Architecture ». *ISPRS Journal of Photogrammetry and Remote Sensing* 158 [décembre]: 11-22. <https://doi.org/10.1016/j.isprsjprs.2019.09.016>.
- Interdonato, Roberto, Dino Ienco, Raffaele Gaetano, et Kenji Ose. 2019. « DuPLO: A DUal View Point Deep Learning Architecture for Time Series ClassificatiON ». *ISPRS Journal of Photogrammetry and Remote Sensing* 149 [mars]: 91-104. <https://doi.org/10.1016/j.isprsjprs.2019.01.011>.
- Janssen, Lucas L. F. 2004. *Principles of Remote Sensing: An Introductory Textbok*. Enschede: ITC.
- Karasiak, N., D. Sheeren, M. Fauvel, J. Willm, J.-F. Dejoux, et C. Monteil. 2017. « Mapping tree species of forests in southwest France using Sentinel-2 image time series ». In *2017 9th International Workshop on the*

Analysis of Multitemporal Remote Sensing Images (MultiTemp), 1-4. Brugge, Belgium: IEEE.
<https://doi.org/10.1109/Multi-Temp.2017.8035215>.

Le Goff, Héloïse, Louis De Grandpré, Daniel Kneeshaw, et Pierre Bernier. 2010. « Sustainable Management of Old-Growth Boreal Forests: Myths, Possible Solutions and Challenges ». *The Forestry Chronicle* 86 [1]: 70-76.
<https://doi.org/10.5558/tfc86070-1>.

Leeuwen, Martin van, et Maarten Nieuwenhuis. 2010. « Retrieval of Forest Structural Parameters Using LiDAR Remote Sensing ». *European Journal of Forest Research* 129 [4]: 749-70. <https://doi.org/10.1007/s10342-010-0381-4>.

Lim, Kevin, Paul Treitz, Michael Wulder, Benoît St-Onge, et Martin Flood. 2003. « LiDAR Remote Sensing of Forest Structure ». *Progress in Physical Geography: Earth and Environment* 27 [1]: 88-106.
<https://doi.org/10.1191/0309133303pp360ra>.

Luo, Shezhou, Cheng Wang, Xiaohuan Xi, Sheng Nie, Xieyu Fan, Hanyue Chen, Xuebo Yang, Dailiang Peng, Yi Lin, et Guoqing Zhou. 2019. « Combining Hyperspectral Imagery and LiDAR Pseudo-Waveform for Predicting Crop LAI, Canopy Height and above-Ground Biomass ». *Ecological Indicators* 102 [juillet]: 801-12.
<https://doi.org/10.1016/j.ecolind.2019.03.011>.

Malfait, Jean-Jacques, et Jean-Christophe Martin. 2017. « Identification des filières économiques à partir des modèles entrées-sorties : l'exemple de la filière bois en France ». *Revue d'économie industrielle*, n° 158 (juin): 69-100. <https://doi.org/10.4000/rei.6564>.

Monnet, Jean-Matthieu. 2015. « ESTIMATION DE PARAMÈTRES FORESTIERS PAR DONNÉES LIDAR AÉROPORTÉ ET IMAGERIE SATELLITAIRE RAPIDEYE : ÉTUDE DE SENSIBILITÉ ». *Revue Française de Photogrammétrie et de Télédétection* 1 [211-212]: 71-79. <https://doi.org/10.52638/rfpt.2015.544>.

Morgan, Jessica L., Sarah E. Gergel, et Nicholas C. Coops. 2010. « Aerial Photography: A Rapidly Evolving Tool for Ecological Management ». *BioScience* 60 [1]: 47-59. <https://doi.org/10.1525/bio.2010.60.1.9>.

Narine, Lana L., Sorin C. Popescu, et Lonesome Malambo. 2019. « Synergy of ICESat-2 and Landsat for Mapping Forest Aboveground Biomass with Deep Learning ». *Remote Sensing* 11 [12]: 1503.
<https://doi.org/10.3390/rs11121503>.

Pan, Y., R. A. Birdsey, J. Fang, R. Houghton, P. E. Kauppi, W. A. Kurz, O. L. Phillips, et al. 2011. « A Large and Persistent Carbon Sink in the World's Forests ». *Science* 333 [6045]: 988-93.
<https://doi.org/10.1126/science.1201609>.

Silveira, Hilton Luís Ferraz da, Lênio Soares Galvão, Ieda Del'Arco Sanches, Iedo Bezerra de Sá, et Tatiana Ayako Taura. 2018. « Use of MSI/Sentinel-2 and Airborne LiDAR Data for Mapping Vegetation and Studying the Relationships with Soil Attributes in the Brazilian Semi-Arid Region ». *International Journal of Applied Earth Observation and Geoinformation* 73 [décembre]: 179-90. <https://doi.org/10.1016/j.jag.2018.06.016>.

Simberloff, Daniel. 1999. « The Role of Science in the Preservation of Forest Biodiversity ». *Forest Ecology and Management* 115 [2-3]: 101-11. [https://doi.org/10.1016/S0378-1127\(98\)00391-0](https://doi.org/10.1016/S0378-1127(98)00391-0).

Simon, J. W., et W. Bode. 2012. « The Citizens Forest Model: Climate Change, Preservation of Natural Resources and Forest Ethics ». In *Climate Change and Sustainable Development*, édité par Thomas Potthast et Simon Meisch, 137-41. Wageningen: Wageningen Academic Publishers. https://doi.org/10.3920/978-90-8686-753-0_18.

Thenkabail, Prasad Srinivasa, éd. 2016. *Land resources monitoring, modeling, and mapping with remote sensing*. Boca Raton, Fla: Taylor & Francis.

Véga, Cédric, Jean-Pierre Renaud, Sylvie Durrieu, et Marc Bouvier. 2016. « On the Interest of Penetration Depth, Canopy Area and Volume Metrics to Improve Lidar-Based Models of Forest Parameters ». *Remote Sensing of Environment* 175 [mars]: 32-42. <https://doi.org/10.1016/j.rse.2015.12.039>.

Wang, Dezhi, Bo Wan, Jing Liu, Yanjun Su, Qinghua Guo, Penghua Qiu, et Xincan Wu. 2020. « Estimating Aboveground Biomass of the Mangrove Forests on Northeast Hainan Island in China Using an Upscaling Method from Field Plots, UAV-LiDAR Data and Sentinel-2 Imagery ». *International Journal of Applied Earth Observation and Geoinformation* 85 [mars]: 101986. <https://doi.org/10.1016/j.jag.2019.101986>.

Wu, Zhuoting, Dennis Dye, John Vogel, et Barry Middleton. 2016. « Estimating Forest and Woodland Aboveground Biomass Using Active and Passive Remote Sensing ». *Photogrammetric Engineering & Remote Sensing* 82 [4]: 271-81. <https://doi.org/10.14358/PERS.82.4.271>.

Zhang, Linjing, Zhenfeng Shao, Jianchen Liu, et Qimin Cheng. 2019. « Deep Learning Based Retrieval of Forest Aboveground Biomass from Combined LiDAR and Landsat 8 Data ». *Remote Sensing* 11 [12]: 1459. <https://doi.org/10.3390/rs11121459>.

Zhong, Linsheng, Jinyang Deng, et Baohui Xiang. 2008. « Tourism Development and the Tourism Area Life-Cycle Model: A Case Study of Zhangjiajie National Forest Park, China ». *Tourism Management* 29 [5]: 841-56. <https://doi.org/10.1016/j.tourman.2007.10.002>.