



HAL
open science

Traitement de données LiDAR pour la modélisation d'indicateurs de biodiversité forestière

Sylvain Jourdana

► **To cite this version:**

Sylvain Jourdana. Traitement de données LiDAR pour la modélisation d'indicateurs de biodiversité forestière. Biodiversité et Ecologie. 2017. hal-02789254

HAL Id: hal-02789254

<https://hal.inrae.fr/hal-02789254v1>

Submitted on 5 Jun 2020

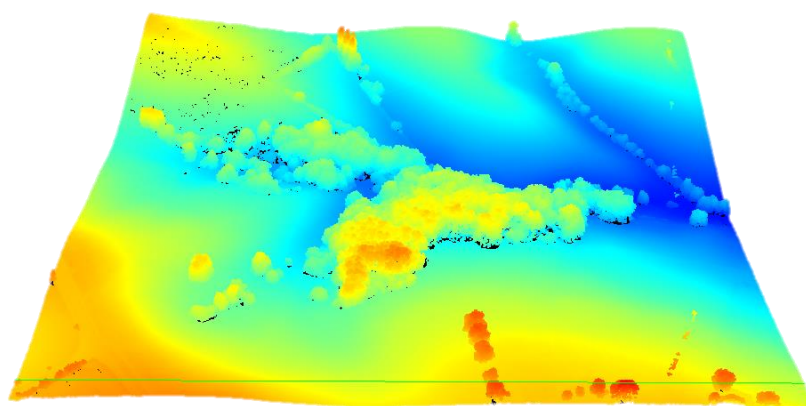
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

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.

Rapport de stage

Cycle : ING3

Traitement de données LiDAR pour la modélisation
d'indicateurs de biodiversité forestière



Sylvain JOURDANA

le 27 septembre 2017

Non confidentiel Confidentiel IGN Confidentiel Industrie jusqu'au

Commanditaire :

INRA UMR DYNAFOR

Chemin de Borde Rouge CS 52627

31326 Castanet Tolosan cedex

Encadrement du stage :

David SHEEREN, UMR DYNAFOR, maître de stage

Ana-Maria OLTEANU-RAIMOND, COGIT, ENSG/IGN, rapporteur principal

Mathieu FAUVEL, UMR DYNAFOR, co-encadrant

Responsable pédagogique du cycle :

Serge BOTTON, cycle ingénieur

Stage du 03/04/17 au 15/09/17

Diffusion Web : Internet Intranet ENSG

Nombre de pages : 176 dont 94 d'annexes

Remerciements

Je tiens tout d'abord à remercier le directeur de l'unité de recherche, Marc Deconchat, pour m'avoir accueilli au sein du laboratoire.

Je suis ensuite reconnaissant envers l'ensemble des membres de l'unité Dynafor, l'INRA et l'ENSAT, pour leur accueil chaleureux. Je les remercie autant pour leur aide et conseils avisés dans mes travaux de recherche que pour leur générosité dans la vie du laboratoire. Ils ont su se montrer disponibles lorsque j'ai pu les solliciter afin de m'aiguiller dans mon étude. Ils m'ont apporté des conseils précieux notamment en matière d'écologie du paysage et de statistiques au travers des différents échanges. Je remercie le personnel administratif qui a su m'épauler dans mes démarches.

Je remercie en particulier mon maître de stage David Sheeren pour l'intérêt qu'il a porté à mon travail. Il a su être présent lorsque j'avais besoin de ses conseils et de son expérience. Mes remerciements s'adressent aussi à mon co-encadrant Mathieu Fauvel qui a répondu à mes sollicitations et aiguillé mes travaux. Leur passion de la recherche et leur rigueur dans le travail sont appréciables et ont été fortes utiles pour la réalisation de mes propres travaux.

Je tiens aussi à remercier chaleureusement le duo Richard Auriol, l'informaticien chevronné, et Jérôme Rigobert, le volleyeur aguerris, pour m'avoir accepté dans leur équipe élite INRA de pétanque. Leur générosité, leur disponibilité, leur sportivité sont régulièrement remarquées dans la vie de l'unité. Je remercie Vincent Thierion et Pierre-Alexis Herrault pour m'avoir accompagné sur le terrain et permis de mieux comprendre les données à ma disposition. Un grand merci également envers l'ensemble des stagiaires, doctorants et post-doctorants du laboratoire dont les échanges durant et en-dehors des heures de travail ont été enrichissantes.

Je remercie aussi l'ensemble du corps enseignant de l'Ecole Nationale des Sciences Géographiques et en particulier Ana-Maria Olteanu-Raimond ma référente lors de ce stage pour sa disponibilité, ses conseils et son suivi régulier.

Je ne saurais terminer sans penser à remercier l'ensemble des personnes de mon entourage qui ont été présentes à mes côtés, contribuant à la réussite de ce stage ; ainsi que l'ensemble des personnes ayant participé à la correction de ce rapport.

Résumé

Les chercheurs en écologie du paysage essaient de comprendre l'effet de la structure et de la composition des habitats sur les espèces. La technologie LiDAR est encore peu utilisée dans ce domaine mais fournit des données précises sur de grandes zones. Nous cherchons ici à évaluer l'apport des données LiDAR en milieu forestier. La but est de voir si des métriques LiDAR pertinentes peuvent être extraites afin d'expliquer et prédire des indicateurs de biodiversité.

Les résultats obtenus à partir de modèles linéaires généralisés (GLM) permettent d'expliquer entre 14 et 63% des variables de biodiversité étudiées (Indice de Biodiversité Potentielle, taux de prédation des défoliateurs, indice de consommation du chevreuil, richesse de la strate herbacée, richesse de différentes catégories d'oiseaux) à partir d'une combinaison d'une quinzaine de variables LiDAR.

Dans un second temps, nous explorons l'apport de données LiDAR pour extraire automatiquement les cimes et couronnes des arbres. L'analyse est effectuée par extraction des maxima locaux ou par l'utilisation de la méthode de ligne de partage des eaux, sur le modèle numérique de hauteur issu des données LiDAR. Les algorithmes ont été testés sur une zone d'étude simplifiée présentant des arbres alignés avec un écart assez fixe. Il en ressort que la détection des cimes d'arbres est, dans ce cas d'étude, précise après un filtrage gaussien 3*3 du raster analysé.

Les résultats obtenus demandent à être confirmés en raison du faible nombre d'observations. Cette étude ouvre toutefois des perspectives intéressantes sur l'apport du LiDAR en écologie.

Mots-clés : biodiversité, cartographie, écologie, explication, foresterie, FUSION/LDV, GLM, LAStools, LiDAR, modèle numérique de hauteur, modélisation statistique, nuage de points, prédiction, R, segmentation, services écosystémiques, spatialisation

Abstract

Landscape ecology researchers are trying to understand the effect of habitat structure and composition on species. LiDAR technology is still not widely used in this domain but provides accurate data over large areas. In this report, we try to assess the contribution of LiDAR data in forest environments. The aim is to see if relevant LiDAR metrics can be extracted to explain and predict biodiversity indicators.

The results obtained from generalized linear models (GLM) allow to explain between 14 and 63% of the biodiversity variables studied (Potential Biodiversity Index, defoliator predation rate, deer consumption index, richness of the herbaceous stratum, richness of different categories of birds) from a combination of about fifteen LiDAR variables.

In a second step, we explore the contribution of LiDAR data to automatically extract treetops and crowns from trees. The analysis is carried out by extracting local maxima or by using the watershed method on the digital height model derived from LiDAR data. The algorithms were tested on a simplified study area with trees aligned with a rather fixed gap. It appears that the detection of tree tops is, in this case of study, accurate after a 3*3 Gaussian filtering of the raster analyzed.

The results obtained need to be confirmed due to the low number of observations. However, this study opens interesting perspectives on LiDAR's contribution to ecology.

Keywords: biodiversity, digital elevation model, ecology, ecosystem services, explanation, forestry, FUSION/LDV, GLM, LAsTools, LiDAR, mapping, prediction, scatter, segmentation, statistical modeling, R

Sommaire

Introduction	8
1. Présentation du contexte	9
2. LiDAR et écologie : état de l'art.....	12
3. Matériels	24
4. Modélisation de variables écologiques à partir de variables LiDAR.....	30
5. Extraction automatique des cimes et couronnes des arbres.....	56
Conclusion	67

INTRODUCTION

Dans un contexte de diminution de la biodiversité, la nécessité de posséder des données spatialisées devient urgente afin d'alimenter des politiques de conservation. La télédétection permet à la fois d'obtenir des données sur une grande étendue et à forte résolution spatiale. Elle représente un apport potentiellement efficace et non négligeable pour les décideurs politiques.

Ce stage s'inscrit dans le cadre d'un projet d'estimation d'indicateurs de biodiversité grâce à l'apport des données de télédétection. L'objectif du projet est de modéliser ces indicateurs afin de les prédire et les spatialiser dans de vastes paysages agricoles et forestiers pour, à terme, créer une carte des indicateurs de biodiversité à l'échelle nationale. Ce travail se base sur des levés LiDAR issus d'une mission aérienne effectuée dans un site de recherches écologiques à long terme du laboratoire Dynafor, au sud-ouest de Toulouse. A cela s'ajoutent le niveau de plusieurs facteurs mesurés sur le terrain, servant à caractériser la forêt et sa biodiversité (par exemple l'indice d'abrutissement lié au chevreuil, le taux de prédation des chenilles, l'Indice de Biodiversité Potentielle).

Cette étude concerne plus précisément le traitement et l'apport des données LiDAR dans le cadre de la prédiction d'indicateurs de biodiversité en milieu forestier. Il s'organise autour de quatre objectifs principaux : (i) effectuer un état de l'art sur l'usage des données LiDAR pour des applications en écologie (ii) extraire du nuage de points des métriques LiDAR qualifiant la forêt (par exemple la hauteur des arbres, la densité de canopée) (iii) modéliser statistiquement les variables écologiques à partir des données de télédétection afin de prédire et spatialiser les indicateurs de biodiversité étudiés (iv) tester des méthodes permettant d'individualiser automatiquement des arbres.

1. PRESENTATION DU CONTEXTE

I. Organisme d'accueil

Le laboratoire Dynafor dans lequel j'ai effectué mon stage est une unité mixte de recherche (UMR) interdisciplinaire (schéma d'organisation en figure 1). Le laboratoire étudie l'écologie des paysages agri-forestiers afin de faire face aux changements globaux actuels. Il est composé de 8 chercheurs, 13 enseignants-chercheurs, 16 ITA, 14 doctorants, 15 stagiaires, complété par une vingtaine de doctorants et stagiaires chaque année.

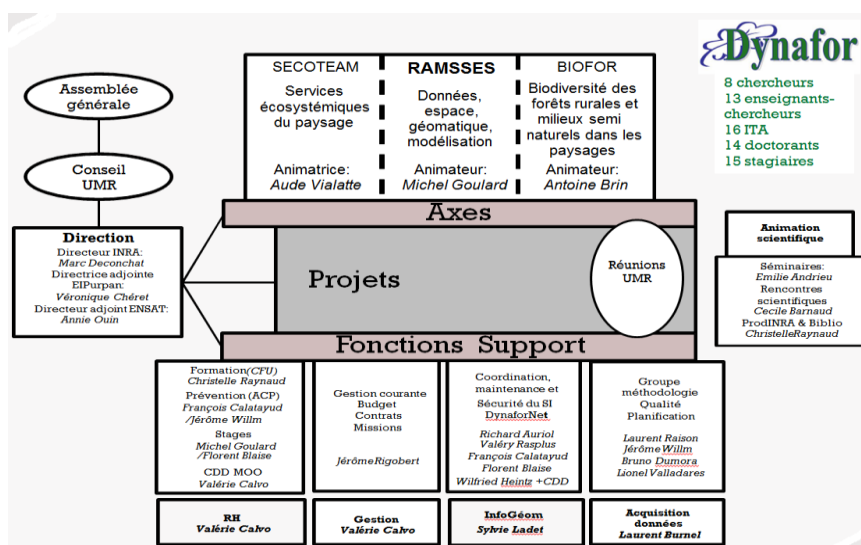


Figure 1 Schéma d'organisation de l'UMR.

Les études sont menées autour de trois axes (Figure 2) : BIOFOR (Biodiversité des forêts rurales et des milieux semi naturels dans les paysages), SECOTEAM (Services écosystémiques du paysage) et RAMSSES (Recherche en Analyse et Modélisation de Systèmes Socio-Ecologiques Spatialisées). Le premier s'intéresse à la gestion durable des forêts tout en étudiant les interactions de ces dernières avec les espaces agricoles. La biodiversité de ces paysages est ainsi évaluée et suivie tout comme l'influence des pratiques forestières sur sa conservation. SECOTEAM est lui tourné vers les bénéfices obtenus par les humains grâce aux écosystèmes (notion de service écosystémique qui sera développée plus loin) et de l'ingénierie écologique par la gestion des habitats. Enfin, le dernier axe, RAMSSES, est celui du croisement entre la géomatique, la télédétection et les statistiques. Il a pour but de construire des méthodologies visant à extraire, analyser et modéliser les paysages et leur dynamique. L'ensemble des axes interagissent pour obtenir une vision intégrée du paysage, dans ses dimensions spatiales,

écologiques et socio-techniques. Ce stage s'inscrit donc dans l'axe RAMSSES ainsi que dans un des projets du laboratoire à savoir le projet MUESLI.

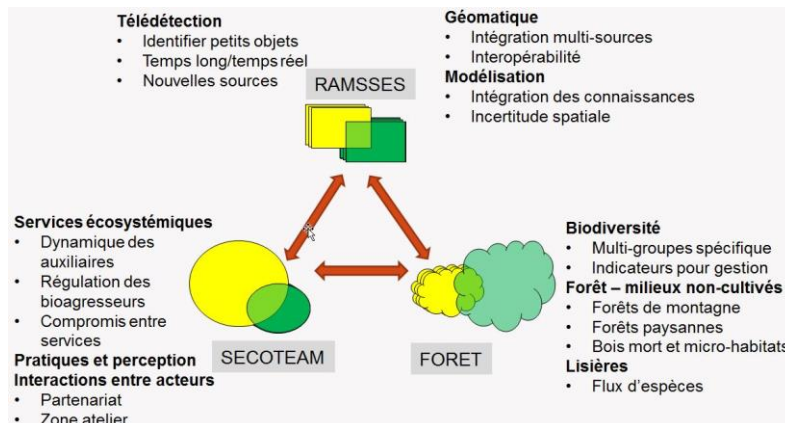


Figure 2 Axes de recherche de Dynafor, source : site internet du laboratoire.

II. Projet interdisciplinaire MUESLI

1) Estimation d'indicateurs de biodiversité dans des paysages agriforestiers

Le projet MUESLI (MULTiscale mapping of Ecosystem Services by very high spatial resolution hyperspectraL and LiDAR remote sensing Imagery) a débuté en 2015. Il s'agit d'un projet d'estimation d'indicateurs de biodiversité à l'aide de données à haute résolution spatiale : des images hyperspectrales et un nuage de points LiDAR. Actuellement, seuls des relevés terrains permettent d'évaluer ces indicateurs or il paraît inconcevable d'utiliser de telles techniques sur de larges échelles. L'objectif du projet est donc de modéliser ces indicateurs afin de les prédire et les spatialiser dans de vastes paysages agricoles et forestiers. Cette démarche s'éloigne des approches courantes qui consistent à définir différentes classes d'occupation des sols permettant de caractériser des objets spatiaux. La durée du projet est de 3 ans et se décompose en trois étapes principales : l'acquisition des données (relevés terrains et levés hyperspectraux et LiDAR), l'état de l'art de la méthode et la construction de la base de données et enfin l'implémentation de la représentation des paysages.

2) Apport potentiel de la télédétection

Comme nous l'avons évoqué précédemment, ce projet ne saurait exister sans des données de télédétection. Cette technologie permet d'obtenir des informations riches et complémentaires aux relevés terrains voire permettant de remplacer ces relevés lors d'une automatisation du processus.

D'un côté, les images hyperspectrales pourraient permettre de décrire plus précisément l'hétérogénéité des paysages grâce aux différentes signatures spectrales contenues dans l'image. D'un autre côté, les données LiDAR fournissent des renseignements précis sur la topographie des lieux et sur la structure verticale des objets notamment en milieu forestier. Néanmoins, cet apport de la télédétection reste conditionné au fait que cette source de données reste difficile à traiter et impose des traitements informatiques lourds.

III. Problématique et objectifs du stage

Mon rôle dans ce projet concerne le traitement des données LiDAR acquises pour le projet MUESLI ; l'objectif étant d'évaluer leur apport dans le cadre de la prédiction d'indicateurs de biodiversité en forêt. Il s'agit ainsi d'extraire du nuage de points des métriques LiDAR qualifiant la forêt (par exemple la hauteur des arbres, leur diamètre moyen, la densité de canopée etc.) à l'échelle d'un bois, d'un plot voire à l'échelle de l'arbre. Ces métriques vont décrire la structure verticale des paysages. Ensuite, il convient de faire le lien entre les variables LiDAR extraites et les variables paysagères connues (relevées sur le terrain) afin d'évaluer quelles variables sont pertinentes dans l'explication et la prédiction des indicateurs de biodiversité étudiés. Ces indicateurs regroupent à la fois certains services écosystémiques (notion que nous définirons par la suite) mesurés dans les forêts, des indices de richesse de différentes communautés d'oiseaux ainsi qu'un indice de biodiversité potentielle en forêt (IBP). Un objectif secondaire consiste à explorer l'apport de données LiDAR pour extraire automatiquement les cimes et couronnes des arbres. Cette extraction peut alors aider à obtenir de nouvelles variables LiDAR, à l'échelle de l'arbre, pouvant être pertinentes pour l'estimation des indicateurs de biodiversité présentés précédemment.

IV. Organisation du travail

La première partie du travail a consisté en un état de l'art sur la technologie LiDAR et ses applications en écologie notamment dans le domaine de la foresterie. Une fois ces recherches effectuées, deux axes de recherche se sont dégagés : la modélisation de variables écologiques à partir de variables LiDAR et l'extraction automatique des cimes et couronnes des arbres. Cette deuxième partie étant vouée à générer de nouvelles variables LiDAR pour améliorer les modélisations construites dans la première. Comme nous pouvons le voir dans le diagramme de GANTT en annexe 1, une plus grande partie de ce stage a été consacrée à la modélisation. Les travaux sur l'extraction de variables à l'échelle des arbres constituent donc les prémices pour des recherches futures.

2. LIDAR ET ECOLOGIE : ETAT DE L'ART

1. Rappel du principe d'un levé LiDAR aéroporté

La mesure LiDAR (Light Detection And Ranging), apparue dans les années 1990, sert à l'acquisition de données altimétriques. C'est une technologie de télédétection active reposant sur l'émission et la réception d'une impulsion laser. Elle permet d'obtenir la distance entre le récepteur et la cible en connaissant la vitesse de déplacement de l'impulsion laser et son temps de parcours (Baltsavias 1999). Cette technique fonctionne nuit et jour mais est dépendante de la météo (modification de l'onde laser réfléchi au contact des particules). Dans le cadre de notre étude, les relevés LiDAR sont aériens : l'avion est alors muni d'un système de positionnement tridimensionnel (GPS et centrale inertielle), d'un distance-mètre laser (capteur LiDAR) et d'un mécanisme de déviation du faisceau (laser à balayage). Le système émet une impulsion laser (généralement dans les longueurs d'onde du proche infra-rouge à 1064 nm) qui va parcourir l'atmosphère avant d'interagir avec les objets au sol. Le signal retour est enregistré par le capteur, en intensité, et transformé ensuite en signal numérique. Cet enregistrement peut correspondre à un ensemble d'impulsions retour (cas du LiDAR à retour discret) ou à un signal continu (LiDAR à retour d'onde complète, ROC). Le signal est alors traité pour fournir un nuage de points 3D, géoréférencés, et plus ou moins dense suivant la cadence d'émission du laser. Ce signal dépend de différents paramètres comme les propriétés radiométriques et géométriques des objets interceptés ou bien encore du diamètre du faisceau émis. Son traitement permet d'extraire différents échos à l'aide des modes présents sur l'onde de retour. On distingue les échos simples (un retour unique, i.e. un seul mode) des échos multiples (plusieurs rétrodiffusions enregistrées pour une même impulsion laser). Des échos simples s'enregistrent par exemple pour des points au sol (cas des routes) tandis que des données de végétation peuvent renvoyer jusqu'à neuf échos. Les bordures de bâtiments peuvent présenter un phénomène de « double écho » c'est-à-dire que deux retours vont être renvoyés pour une même émission à cause du rebond de l'onde sur les parois (Mallet cours 2017).

Ce nuage de points (exemple en figure 3) peut être caractérisé par une série de métriques décrivant sa distribution dans l'espace. Il est également possible d'en dériver plusieurs produits raster : un Modèle Numérique de Terrain (MNT), un Modèle Numérique de Surface (MNS) incluant les éléments du sursol et un Modèle Numérique de Hauteur (MNH = MNS - MNT, Figure 3).

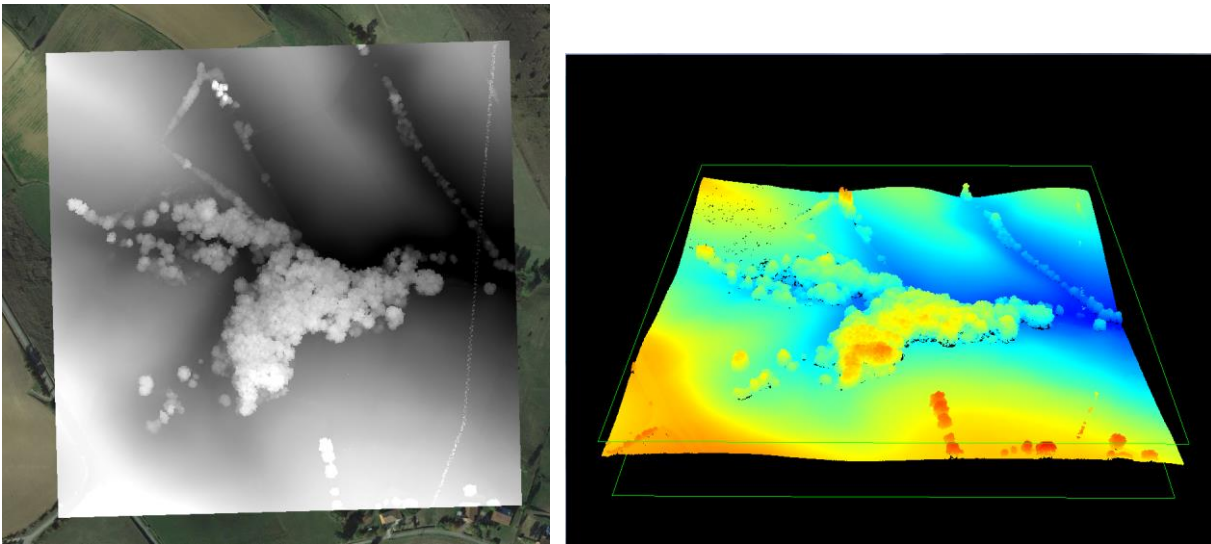


Figure 3 MNH raster d'un bois (à gauche) et nuage de points LiDAR colorisé suivant l'altitude (à droite).

Lors de l'achat de données LiDAR auprès d'un producteur spécialisé, il est généralement demandé de fournir un nuage de points partiellement classés. Les points correspondant à des retours au sol sont distingués des autres grâce à la nature de l'écho renvoyé (les premiers échos d'une série de plusieurs correspondent forcément au sursol) puis, par analyse du terrain et filtrage spatial. Par la suite, des analyses plus approfondies, souvent manuelles, peuvent permettre d'affiner la classification des autres points (végétation - visible en figure 4, bâtiment etc.).

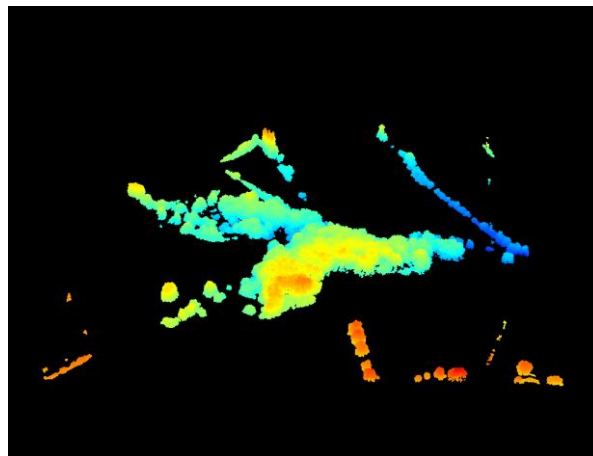


Figure 4 Classe végétation d'un nuage de points LiDAR, colorisé suivant l'altitude des points.

Le format standard de distribution du nuage de points LiDAR est le LAS (LASer file format exchange). Il s'agit d'un format standardisé dont les spécifications¹ sont bien définies, contenant par exemple les coordonnées des points, leur intensité, le numéro du retour, la classification, l'angle de scan etc.

¹ <https://www.asprs.org/committee-general/laser-las-file-format-exchange-activities.html>

II. Usage des données LiDAR pour des questions écologiques

1) Apport du LiDAR en écologie et notamment foresterie

De nombreuses recherches ont été effectuées afin d'utiliser des données LiDAR dans des modèles écologiques. Nous pouvons classer les applications liant LiDAR et écologie en trois catégories principales : (1) la connaissance précise de la topographie des habitats naturels, (2) la mesure de la structure verticale de la végétation et (3) la prédiction d'attributs d'écosystèmes forestiers tels la biomasse (Lefsky et al. 2002). Plus généralement, ces données de télédétection permettent d'obtenir une vision plus précise, en 3D et à grande échelle de la structure des habitats, que ce soit au niveau de la topographie (sol) ou de la structure de la végétation (Gachet 2009) et ses paramètres dérivés (sursol).

Ces données influencent donc les recherches écologiques car la structure 3D des habitats est encore peu prise en compte dans les modèles de distribution d'espèces. Simonson et al. (2014) s'intéressent par exemple à l'estimation de la diversité en espèces animales à partir de relevés LiDAR aérien. Cette estimation se base sur l'importance de la structure de l'habitat pour expliquer la diversité des espèces animales. En effet, des modèles existent pour faire le lien entre cette structure et la diversité. Ils ont ainsi étudié 269 articles publiés entre 1960 et 2013 pour relever les relations entre différentes espèces et des attributs LiDAR, permettant de caractériser la structure de leurs habitats spécifiques.

Les indicateurs de la structure végétale de l'habitat sont divisés en trois parties : ceux relevant de la structure verticale, ceux relevant de la structure horizontale et les autres indicateurs (par exemple la diversité en arbres). Ainsi, des relations empiriques ont été mises en évidence entre, par exemple, la présence de primates en forêt et la hauteur moyenne et maximum de la végétation (Schwarzkopf & Rylands 1989), entre le nombre et la diversité d'occupation du sol et la richesse en reptiles et amphibiens (Rodríguez, Belmontes & Hawkins 2005) ou encore entre la diversité en plantes et les coléoptères dans les broussailles (González-Megías, Gómez & Sánchez-Piñero 2011). Des études ont également été menées sur l'utilisation de mesures LiDAR pour évaluer de telles structures d'habitats forestiers par exemple la hauteur moyenne et maximale (Parker & Russ 2004), la densité d'arbres (Lefsky et al. 1999) et les variations de hauteurs (Donoghue et al. 2007). Pour cela, différentes variables sont extraites des données LiDAR afin de caractériser l'habitat de la zone levée. Ces variables seront détaillées dans la section suivante, il s'agit par exemple du volume et de la hauteur de végétation.

T Vierling et al. (2008) ont aussi auparavant cherché des applications de la technologie LiDAR pour la gestion et la conservation des espèces. Ce choix est réalisé afin de ne pas être limité spatialement comme le permettent les relevés terrain ni verticalement comme le conçoit la télédétection par analyse d'images aériennes et ainsi améliorer les modèles utilisés dans le domaine des relations habitats-espèces. La zone d'intérêt peut alors être cartographiée suivant les variables LiDAR afin de construire des modèles statistiques optimaux permettant de prédire la présence ou a contrario l'absence d'espèces grâce à la connaissance des caractéristiques de l'habitat.

Mais les données LiDAR peuvent aussi être utilisées de manière exploratoire afin de chercher les caractéristiques de l'habitat d'une espèce à partir de distributions spatiales connues. Dans chaque cas, les relevés terrains permettent d'évaluer les modèles construits afin de les valider.

Outre l'estimation de la diversité en espèces, les données LiDAR sont souvent utilisées dans le domaine de la foresterie. En effet, de nombreux paramètres utiles dans ce domaine peuvent être partiellement évalués à partir de données LiDAR. Zhao et al. (2011) cherchent ainsi à relier les caractéristiques de la canopée aux données LiDAR à l'aide de l'apprentissage automatique tandis que A. Wulder et al. (2012) estiment des variables dendométriques² par utilisation conjointe de données LiDAR et de relations allométriques (voir section III. 2)). Il s'agit d'estimer des paramètres biophysiques tels la hauteur des arbres, la taille de la couronne, son volume, le diamètre à hauteur de poitrine, la densité de canopée, la biomasse. L'intérêt d'utiliser cette technologie repose notamment sur la précision des relevés et celle des estimations obtenues (en plus de la possibilité de spatialiser ces caractéristiques sur de grandes étendues). Certaines études plus récentes s'orientent aussi sur la détection du bois mort, notamment en différenciant l'intensité LIDAR des arbres vivants de celle des arbres morts (Wing et al. 2015). Ces études sont intéressantes dans le cadre de ce projet car l'un des indicateurs à expliquer (IBP) prend en compte la présence de bois mort. Néanmoins elles sont peu nombreuses et concernent bien souvent des écosystèmes forestiers très atteints (par exemple présence d'insectes ravageurs).

Les différentes études évoquées sont souvent menées à différentes échelles : celle de la « placette », et celle de « l'arbre ». Dans l'approche « placette », des modèles empiriques sont calibrés à l'échelle des placettes en liant la variable dendrométrique d'intérêt relevée sur le terrain aux données LiDAR extraites à cette échelle. La zone d'étude est dans le même temps divisée en une grille dont les cellules sont d'une taille approximant celle des placettes (une cellule = une unité d'échantillonnage). Les valeurs des variables LiDAR sont alors extraites pour chaque cellule, puis,

² Qui se rapportent aux caractéristiques physiques quantifiables des arbres (e.g. hauteur, volume, diamètre).

les modèles calibrés sont appliqués à l'ensemble des cellules de la grille. On obtient ainsi la valeur prédite de l'attribut étudié pour chaque cellule de la zone d'intérêt (voir figure 5 décrivant l'approche par zone d'après White et al. 2013). Nous pouvons ainsi obtenir des valeurs prédites agglomérées à l'échelle d'un peuplement qui tient davantage compte des disparités de structures au sein d'un peuplement que si la variable dendrométrique avait directement été prédite à l'échelle du peuplement.

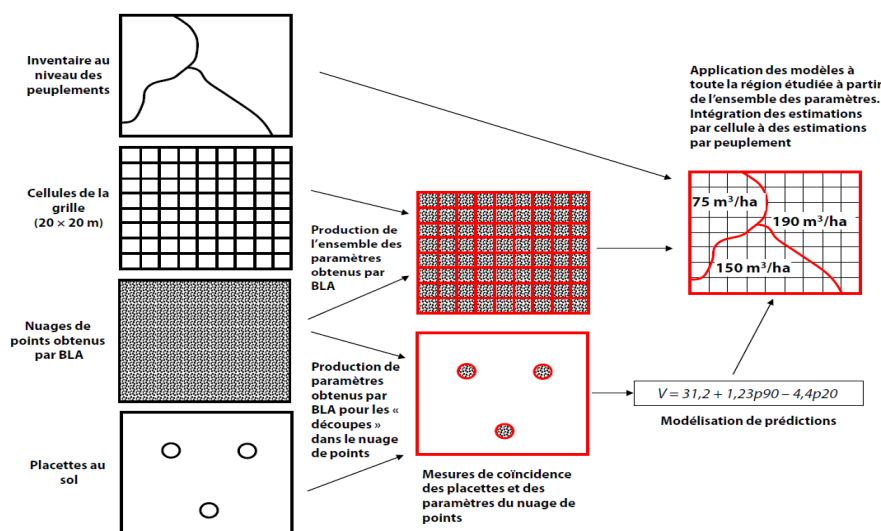


Figure 5 Schéma de l'approche par zone tiré de White et al. (2013) : calibration de modèles prédictifs à partir des relevés terrain des placettes et construction d'une grille dans laquelle les variables LiDAR obtenues par balayage laser aéroporté (BLA) vont être extraites pour chaque cellule de la taille des placettes afin de prédire la variable dendrométrique d'intérêt sur toute la zone d'étude.

L'approche par « arbre » revient quant à elle à segmenter le nuage de points LiDAR afin de connaître les points correspondant à un arbre spécifique (Popescu et al. 2003). Une fois cette segmentation réalisée, les variables LiDAR peuvent être extraites pour chaque objet et ainsi fournir une information à un niveau de détail plus fin. Par exemple, la hauteur arbre par arbre est obtenue et non plus une valeur moyenne à l'échelle du peuplement tout comme le nombre d'arbres par peuplement qui peut permettre d'estimer la surface terrière³ d'un peuplement.

2) Apport du LiDAR pour l'évaluation des services écosystémiques

La partie précédente a passé en revue certains travaux cherchant à mobiliser des données LiDAR pour construire des modèles de distribution d'espèces (spatialisation de la présence ou de la diversité d'une espèce) ou pour prédire des variables de structure dans les écosystèmes

³ Somme des surfaces des sections transversales des troncs à 1.30m de hauteur présents sur une surface donnée.

forestiers. Peu de travaux se sont intéressés à explorer l'intérêt des données LiDAR pour prédire des services écosystémiques, ce que nous cherchons aussi à faire dans le cadre de ce stage.

On entend par services écosystémiques, ou services écologiques, l'ensemble des « biens et services que les hommes peuvent tirer des écosystèmes, directement ou indirectement, pour assurer leur bien-être » (d'après MEA : Millenium Ecosystem Assessment, 2005). Nous comprenons donc que ces services sont en rapport avec l'être humain et les bénéfices qu'il va tirer de la nature. Les écosystèmes, par leurs structures et processus biologiques, assurent certaines fonctions (par exemple le cycle de l'eau) dont on tire bénéfice (par exemple usage de l'eau pour l'irrigation), comme explicité sur la figure 6.

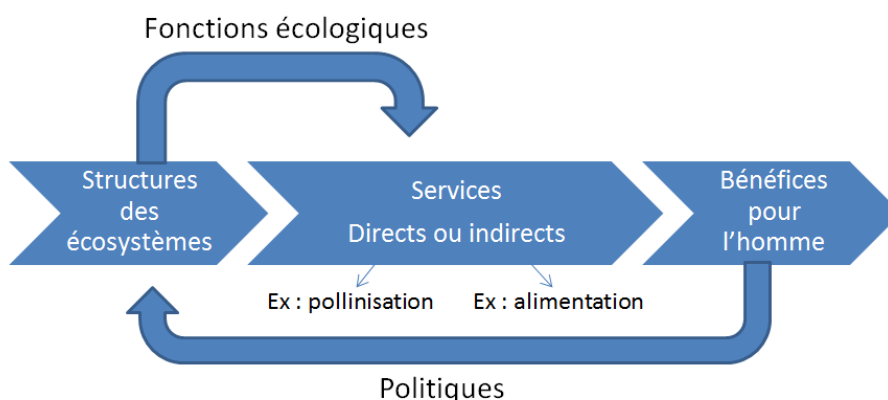


Figure 6 Schéma du lien entre les écosystèmes et l'homme rempli par les services écosystémiques adapté d'UICN France (2012).

Les services écosystémiques peuvent se décomposer en quatre catégories (d'après *Panorama des services écologiques fournis par les milieux naturels de France*, UICN 2012) : les services de support, les services de régulation, les services d'approvisionnement et les services culturels. Les services de support correspondent aux éléments indispensables au fonctionnement des écosystèmes comme par exemple le cycle de l'eau, dont le changement a un impact sur l'homme bien que ce dernier ne l'utilise pas directement. Les services de régulation représentent eux la faculté des écosystèmes à s'autoréguler. Par exemple, ils comprennent le contrôle biologique de ravageurs ou d'espèces nuisibles (régulation par des prédateurs naturels). Ils incluent également la pollinisation, dont le processus assure la reproduction des espèces, mais aussi, la dégradation des déchets par les milieux naturels et la régulation du climat (par le stockage du dioxyde de carbone dans les forêts). Les services d'approvisionnement sont plus évidents et se rapprochent des biens écologiques c'est-à-dire, des ressources fournies par la nature, que ce soit la nourriture (céréales, fruits, produits de la chasse, etc.), l'eau douce ou encore des matières premières (bois, résine, etc.). Enfin, les services culturels sont les services non matériels rendus indirectement par

les écosystèmes. Ils font référence aux effets que peuvent avoir de beaux paysages, une prairie fleurie, ou encore, le bruit d'une rivière, sur nos ressentis, nos valeurs, nos inspirations. Ces services peuvent être à dimension culturelle (aménités) ou de loisirs (tourisme, sport de nature, etc.). Ces services se retrouvent partout sur la planète que ce soit à l'échelle globale ou locale et des politiques ont été mises en place afin de préserver l'ensemble des fonctions écosystémiques répertoriées.

III. Variables LiDAR usuellement rencontrées dans la littérature permettant de décrire le nuage de points

Dans les modèles évoqués précédemment reliant les variables écologiques à prédire et les variables explicatives LiDAR on retrouve souvent des variables servant à caractériser la structure verticale de la végétation notamment la distribution de hauteur d'une placette. Nous allons maintenant décrire ces variables.

1) Extraction de données du nuage de points

Les logiciels destinés à traiter des données LiDAR permettent d'obtenir directement, à partir du nuage de points enregistrés, de nombreuses métriques LiDAR. Nous pouvons obtenir une centaine de variables, plus ou moins semblables les unes des autres. Elles se composent pour l'essentiel de différentes statistiques descriptives sur la hauteur des objets à une étendue spatiale prédéfinie (plot, bois, tuile correspondant à l'échelle des relevés terrain, etc.). Nous pouvons par exemple calculer la hauteur maximale du nuage de points, sa valeur moyenne, médiane, le coefficient de dissymétrie (skewness) ou bien encore, le coefficient d'aplatissement (kurtosis) permettant de qualifier les distributions de hauteur dans la portion de nuage étudiée. De plus les percentiles et bincatiles de hauteur peuvent être calculés. Ces premiers sont expliqués en figure 7. A noter qu'il est possible de calculer ces statistiques pour des tranches de hauteur particulière en ne sélectionnant que les points du nuage s'y trouvant.

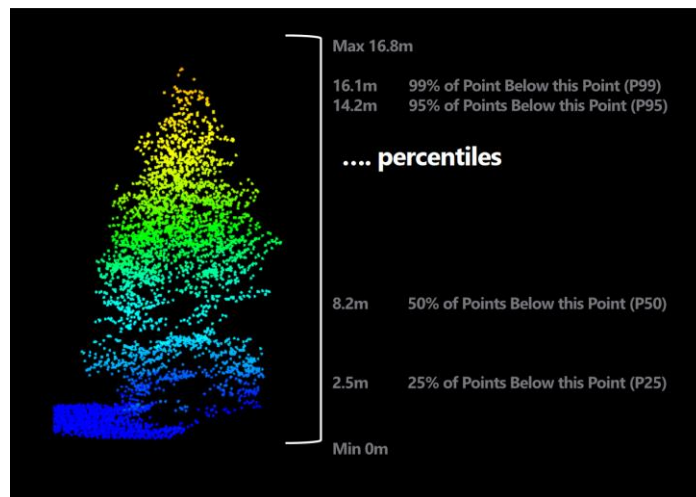


Figure 7 Schéma expliquant le calcul des percentiles tiré de David Herries (2014).

A cela s'ajoute des métriques identiques mais touchant à l'intensité du signal et non plus à la hauteur des objets. Nous disposons par ailleurs du calcul de divers pourcentages attestant du nombre de retour au-dessus d'un seuil de hauteur (par exemple la hauteur moyenne ou le mode) ainsi que du nombre de retours en fonction du numéro de retour (i.e. le nombre de premier retour, le nombre de deuxième, troisième retour etc. comme nous pouvons le voir sur la figure 8).

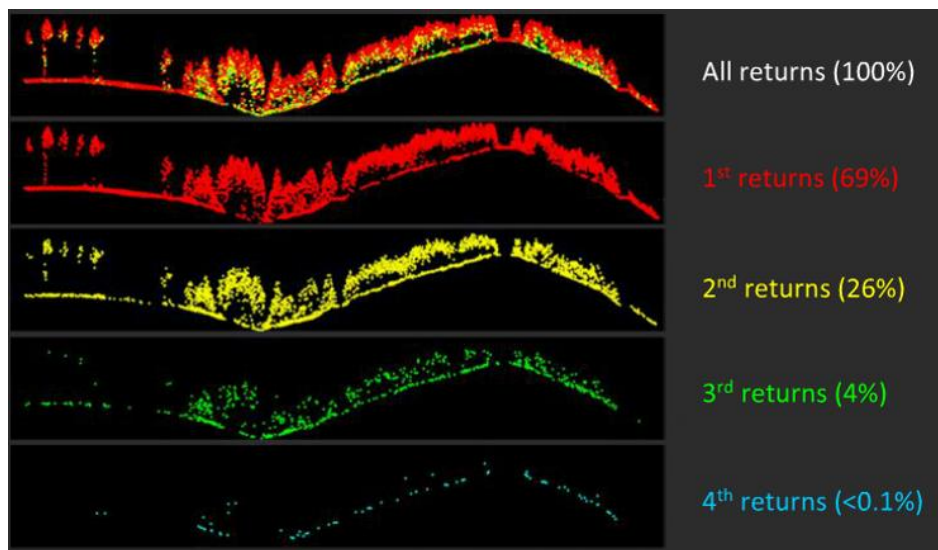


Figure 8 Schéma de différenciation des différents retours LiDAR (proportions approximatives) tiré de l'université d'Humboldt.

Enfin, nous retrouvons des métriques permettant de qualifier la canopée au niveau de sa densité, son relief, sa pénétrabilité etc. Il s'agit notamment de la transparence de canopée (Figure 9), de la densité à partir d'un certain seuil de hauteur et de la couverture de canopée (Figure 10). Il ne s'agit

donc plus ici de regarder les valeurs de hauteur ou d'intensité des points mais bien le nombre de points dans une certaine section.

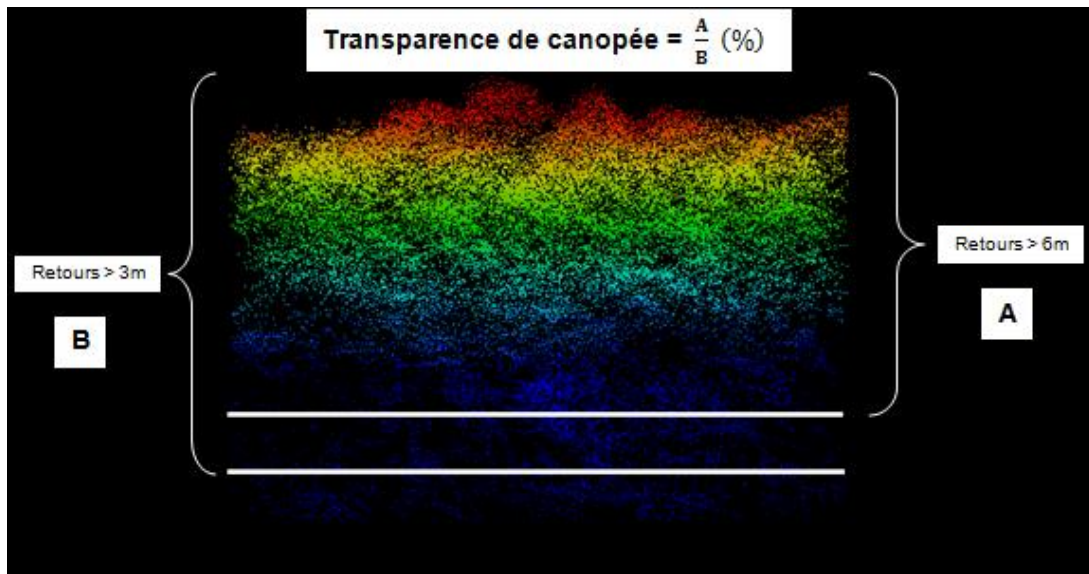


Figure 9 Illustration du calcul de la transparence de canopée adaptée du manuel d'utilisation de FUSION, source : http://forsys.cfr.washington.edu/fusion/FUSION_manual.pdf.

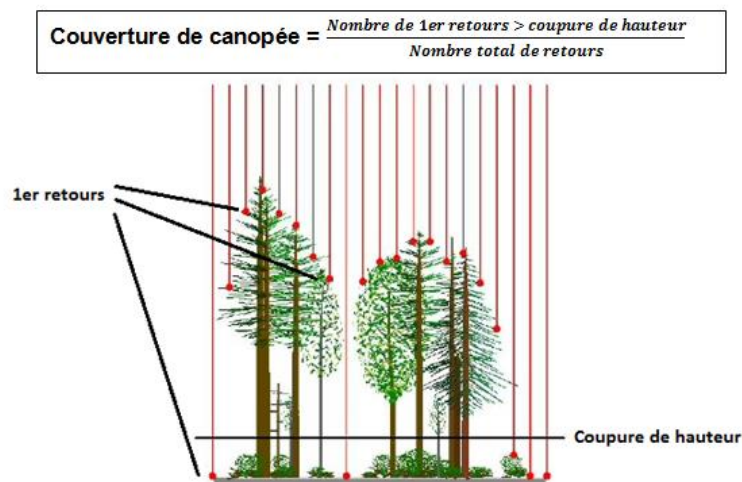


Figure 10 Illustration du calcul de la couverture de canopée adaptée du manuel d'utilisation de FUSION.

Comme nous le verrons par la suite, une fois ces premières métriques extraites, il est possible d'en définir de nouvelles afin d'augmenter le nombre de variables à notre disposition et d'ainsi mieux caractériser le nuage de points LiDAR et donc les bois étudiés.

2) Caractérisation du nuage de points LiDAR à l'échelle de l'arbre

Dans le domaine de la foresterie nous retrouvons aussi beaucoup de mesures effectuées à l'échelle de l'arbre individuel. Certaines de ces mesures sont calculables à partir des données LiDAR, le plus souvent en ayant recourt à des modèles allométriques. Il s'agit de relations généralement non-linéaires existant entre différentes variables qui caractérisent certains traits physiologiques ou morphologiques. A l'aide de mesures de terrain sur de grands effectifs, des modèles mathématiques sont créés afin de décrire la relation entre variables évoluant l'une par rapport à l'autre. Une des équations allométriques les plus connues en foresterie permet de relier la hauteur d'un arbre à son diamètre à hauteur de poitrine c'est-à-dire à 1.30m au-dessus du sol. Il en existe d'autres notamment afin de calculer la couronne d'un arbre ou sa surface terrière à partir de sa hauteur. Le tableau 1, extrait d'un rapport d'étude sur l'utilisation des données LiDAR en milieux forestiers à Mayotte (Lesur 2011), présente certaines de ces relations.

Tableau 1 Variables dérivées de l'analyse de détection des cimes tiré de Lesur (2011).

Diamètre ($d_{1.3m}$ ou dhp)	Fonctions allométriques : $d_{1.3m} = a.Exp(b.h)$ $h = a.Ln(d_{1.3m}) + b$	h : hauteur totale de l'arbre (m), $d_{1.3m}$: diamètre à 1.3 m de hauteur ou hauteur d'homme (dhp), a et b : paramètres du modèle.
Couronne (largeur de la)	Fonction allométrique : $a_0 + a_1.h + a_2.h^2 + a_3.h^3$	h : hauteur totale de l'arbre (m), a_0, a_1, a_2, a_3 paramètres du modèle.
Surface terrière (g ou G)	Pour un arbre (m^2) : $g = \frac{\pi.d_{1.3m}^2}{4}$ Pour un peuplement (m^2/ha) : $G = \sum n_i.g_i$	$d_{1.3m}$: diamètre à 1.3 m de hauteur (m), n_i : nombre d'arbre par hectare, g_i : surface terrière individuel des n_i arbres (m^2).
Volume sur pied (v ou V)	Formules de cubage en « bloc » : Pour un arbre (m^3) : $v = \frac{\pi.d_{1.3m}^2}{4}.h.f = g.f.h$ Pour un peuplement (m^3/ha) : $V = G.F.H$	$d_{1.3m}$: diamètre à 1.3 m de hauteur (m), g, G : surface terrière de l'arbre, du peuplement (m^2/ha), f, F : coefficient de forme moyen de l'arbre, du peuplement, h, H : hauteur totale moyenne de l'arbre, du peuplement (m), v, V : hauteur totale moyenne de l'arbre, du peuplement (m^3/ha).
Biomasse ligneuse aérienne ou (AGB pour Above Ground Biomasse en anglais)	Fonction allométrique (AGB en kg) : $d.Exp(a_0 + a_1.Ln(d_{1.3m}) + a_2.Ln(d_{1.3m})^2 + a_3.Ln(d_{1.3m})^3)$	d : densité spécifique moyenne du bois (g/cm^3), a_0, a_1, a_2, a_3 : paramètres du modèle, $d_{1.3m}$: diamètre à 1.3 m de hauteur en m.

L'information de hauteur étant directement obtenue à partir des données LiDAR, il est intéressant d'extraire les sommets et couronnes des arbres afin de pouvoir calculer ces relations allométriques et prédire de nouvelles variables pour chaque arbre, ce qui n'est pas forcément aisé. A noter que la largeur de couronne peut potentiellement être obtenue à partir de la hauteur de l'arbre, à partir d'un modèle bien calibré. La détection des sommets pourrait donc suffire à obtenir l'emprise individuelle des arbres. Elle fera l'objet de la deuxième partie de ce travail.

IV. Principaux logiciels libres de traitement des données LiDAR

Différents logiciels permettent de traiter les données LiDAR et notamment certains orientés foresterie : ArcGIS, LAStools, LiForest, FUSION/LDV et OPALS. LAStools, FUSION/LDV et OPALS, logiciels, du moins en partie, libres, ont fait l'objet d'une comparaison plus détaillée (Tableau 2). OPALS n'a finalement pas été utilisé dans le cadre de ce stage. En effet, après comparaison il se trouve qu'il ne présentait pas davantage de fonctions utiles pour ce projet dans sa version actuelle que LAStools et FUSION/LDV qui sont faciles à prendre en main et bien documentés. Cependant, ce logiciel est toujours en cours d'implémentation et il pourrait être intéressant de l'utiliser dans des recherches futures notamment pour les modules, en cours de développement, de segmentation et de classification des espèces d'arbres.

Tableau 2 Comparaison de logiciels permettant le traitement des données LiDAR.

	LAStools	FUSION/LDV	OPALS
Prix	Payant en version complète	Gratuit	Gratuit
Système d'exploitation	Windows	Windows	Windows/Linux
Formats d'import	++	++	+++
Formats d'export	++	++	+++
Conversion de format	OUI	OUI	OUI
Filtrage du nuage	+++	++	+++
Métriques	<ul style="list-style-type: none"> - Densité de canopée - Couverture de couvert - Gap fraction - Percentiles + statistiques de hauteur ou d'intensité (moy, min, max, kurtosis, asymétrie, écart-type etc.) 	<ul style="list-style-type: none"> - Maximum de canopée - Modèle de canopée - Percentiles + statistiques de hauteur ou d'intensité (moy, min, max, med, mode, déviation, variance, Skewness, Kurtosis etc.) - Ratio du relief de canopée - Estimation de la couverture de canopée - Divers rapports de rendements au-dessus d'un seuil de hauteur 	<ul style="list-style-type: none"> - Délimitation de la frontière forestière - Délimitation de peuplements de bois homogènes - Gap fraction - Estimation des distributions du diamètre de la tige - Estimation de la biomasse (stock de bois) - Analyse de la structure de la couche forestière Pas encore implémenté : <ul style="list-style-type: none"> - Détermination de la couverture de la couronne - Classification des espèces d'arbres
Segmentation des couronnes	+	+++	-
Contrôle qualité	OUI	OUI	OUI
Avantages complémentaires	Communauté, tutoriels	Manuel, spécialisé foresterie	Orienté recherche

LAStools⁴, développé par Martin Isenburg, présente un ensemble d'outils utilisables en ligne de commandes (aussi disponible dans certains SIG tel QGIS), permettant notamment de classer, normaliser ou bien encore découper des fichiers LiDAR. Certaines fonction sont gratuites alors que d'autres deviennent payantes au-delà d'un certain nombre de points à traiter. Des tutoriaux⁵ et un forum⁶ sont disponibles et facilitent la prise en main du logiciel. Ce logiciel est largement reconnu et utilisé dans la communauté LiDAR.

FUSION/LDV⁷ est lui spécialisé dans l'analyse de données LiDAR pour la foresterie. Il est développé par Robert McGaughey de l'université de Washington (U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station). Il permet notamment d'extraire différentes variables liées à la hauteur des arbres au sein d'un nuage de points (par exemple hauteur moyenne, hauteur maximale etc.). Cet outil est lui aussi très répandu et un des premiers à avoir été diffusé librement.

FugroViewer⁸ permet quant à lui de visualiser des tuiles LiDAR tout en colorisant le nuage de points suivant différents paramètres : élévation, intensité, numéro de retour, classification, identifiant de la source. Il est aussi possible de réaliser des coupes transversales afin d'obtenir une vue de profil ou bien encore de visualiser le nuage en 3D. Ce logiciel est convivial pour la visualisation des données mais ne permet pas d'effectuer des traitements ou calculer des métriques à partir du nuage de points.

⁴ <https://rapidlasso.com/lastools/>

⁵ <https://rapidlasso.com/category/tutorials/>

⁶ <https://groups.google.com/forum/#!forum/lastools>

⁷ <http://forsys.cfr.washington.edu/fusion.html>

⁸ <https://www.fugro.com/about-fugro/our-expertise/technology/fugroviewer>

3. MATERIELS

I. Site d'étude : les coteaux de Gascogne

Lieu d'étude depuis 25 ans dans le cadre des recherches de Dynafor sur les dynamiques de la biodiversité, les coteaux de Gascogne se situent au sud de la Garonne et de Toulouse, dans le Bas Comminges (Figure 11). La région étudiée, plus précisément les communes aux alentours d'Aurignac, présente une agriculture de polyculture et d'élevage. Elle combine à la fois des parcelles cultivées et des prairies mais aussi de nombreux bois, généralement privés, dont la taille peut varier de moins d'un hectare à environ 50 hectares. Ces forêts, principalement constituées de chênes mais aussi d'autres espèces comme le merisier et le charme, sont exploitées pour le bois de chauffage et le bois d'œuvre dans un système de taillis sous futaie ou de taillis simples. Concernant le relief de la zone, nous sommes entre 200 et 400m d'altitude dans une zone vallonnée aux pentes assez fortes notamment dans les zones forestières où l'agriculture est difficile. Le climat est lui soumis à plusieurs influences : atlantique, méditerranéenne et dans une moindre mesure montagnarde, alternant entre étés chauds et secs et hivers doux et humides.

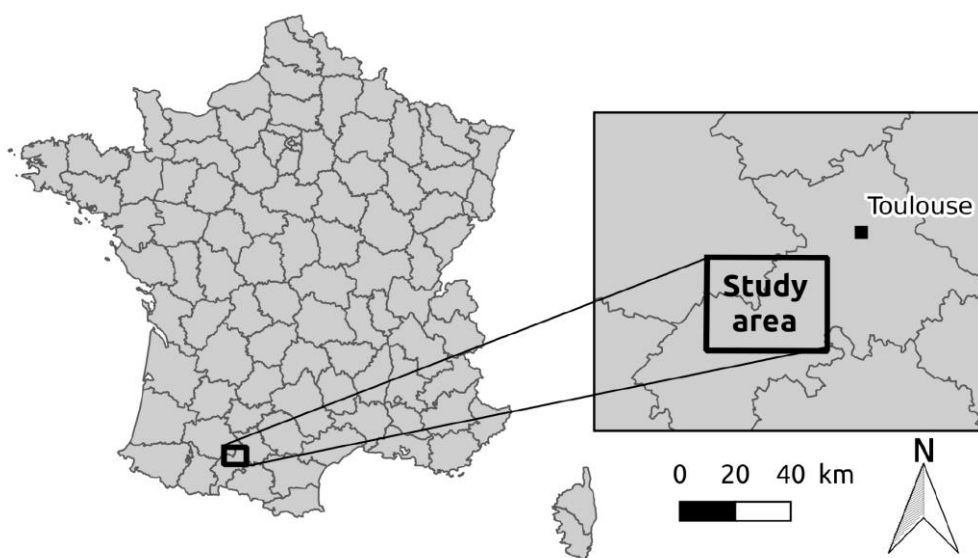


Figure 11 Localisation du site d'étude.

II. Forêts échantillonnées et données écologiques associées

Dans le cadre du projet MUESLI, 30 bois et 30 parcelles agricoles ont été sélectionnés, répondant à certains critères. Pour ce stage, seuls les bois nous intéressent avec le traitement des données

LiDAR associées. C'est pourquoi nous présenterons ici seulement ces entités. La sélection des bois dans le projet répond à deux critères : leur surface doit être comprise entre 1 et 5ha, et ils doivent être « anciens ». Cette ancienneté est jugée en considérant qu'au moins 50% de la surface actuelle du bois doit être présente sur la minute de la carte d'État-major produite autour de 1850. Ces critères sont importants pour les écologues dans le cadre du projet car ils ont un impact sur la biodiversité au sein des écosystèmes forestiers. Nous pouvons ajouter que ces bois sont accessibles facilement et inscrits dans un gradient de paysage boisé variant de 2% à 40% (i.e. la proportion de surface boisée dans un buffer de 1km² autour de ces forêts varie de 2% à 40%). Ces bois sont de type taillis sous futaie et constitués uniquement d'espèces de feuillus à dominante de chênes.

Différents indicateurs écologiques sont associés à ces forêts échantillonnées. Ils ont été calculés par les membres du projet à partir des relevés terrain effectués fin 2015 et début 2016. Il s'agit notamment de :

- Indice de Biodiversité Potentielle (IBP⁹) : indicateur indirecte de la biodiversité d'un peuplement en fonction de sa capacité d'accueil. Le calcul de cet indicateur repose sur dix facteurs : 7 liés au peuplement et à sa gestion et 3 liés au contexte (Annexe 2).

- Abroutissement lié au chevreuil : calculé à partir de l'indice de consommation de la végétation, il traduit la consommation par les ongulés de végétaux ligneux et semi-ligneux sur les placettes relevées.

- Indice de prédation des défoliateurs : pourcentage de fausses chenilles posées dans les arbres qui ont reçu des coups de becs d'oiseaux. Il s'agit d'une mesure directe d'un service de régulation (relation proie-prédateur).

- Richesse spécifique et abondance des oiseaux : calculées à partir de la méthode des points d'écoute avec des relevés effectués en avril (nicheurs précoces actifs) et mai (présence des migrateurs). La richesse exprime le nombre d'espèces d'oiseaux différentes et l'abondance informe sur l'effectif de chaque espèce relevée.

- Richesse botanique : abondance relative (pourcentage de recouvrement de l'espèce sur la surface échantillonnée) et richesse spécifique (nombre d'espèces végétales) pour les strates herbacée et arbustive.

⁹ <http://www.foretpriveefrancaise.com/n/ibp-indice-de-biodiversite-potentielle/n:782>

A cela s'ajoute des données d'IBP issues d'une campagne nationale datant de l'été 2011 (Herrault et al. 2015) dans la zone couverte par les données LiDAR. Ces 45 relevés ont été effectués à l'aide d'un protocole de description rapide (Larrieu and Gonin 2008) dans des plots circulaires d'1ha permettant d'obtenir les variables de structures utiles au calcul de l'IBP.

Les variables évoquées issues des mesures terrain ne sont pas toutes calculées à la même échelle (Tableau 3). En effet, certaines variables ont été calculées à l'échelle du bois alors que d'autres ont été réalisées sur des plots de rayon variable. Nos différentes variables LiDAR seront extraites en tenant compte de cette diversité de tailles d'échantillonnage : plots d'un rayon de 10m (données sur la richesse botanique), plots d'un hectare (données sur l'Indice de Biodiversité Potentielle, relevés à l'échelle de la France), plots de rayons proches de 26m (prédation des défoliateurs), bois (données IBP dans le cadre du projet Muesli, abondance et richesse des oiseaux, données sur l'abrutissement par les chevreuils).

Pour la prédation des chenilles, les plots au sein des bois (de 3 à 4) ont été tirés aléatoirement en évitant leur recouvrement et en faisant attention qu'ils soient totalement inclus dans le bois puis positionnés par GPS. Le rayon maximal du plot est de 26m. Cette distance correspond à la distance maximale à laquelle un arbre est considéré lors de l'acquisition de la surface terrière à l'aide du relascope de Bitterlich¹⁰. En pratique le diamètre dépend de la distance entre le centre du plot étudié et la distance du plus gros arbre présent sur le terrain. Pour cet indicateur, des fausses chenilles sont utilisées comme leurres sur les arbres et les coups de becs donnés par des oiseaux sont relevés (protocole Figure 12). Ils permettent d'estimer un taux de prédation des chenilles qui elles, lorsqu'il ne s'agit pas de leurres, peuvent nuire aux arbres (insecte défoliateur).

¹⁰ Outil permettant d'estimer la surface terrière d'un peuplement en comptabilisant le nombre d'arbres dont le diamètre est supérieur à celui d'une encoche de 1cm de largeur placée à 50cm devant son œil lors d'un tour d'horizon (Bitterlich 1984).

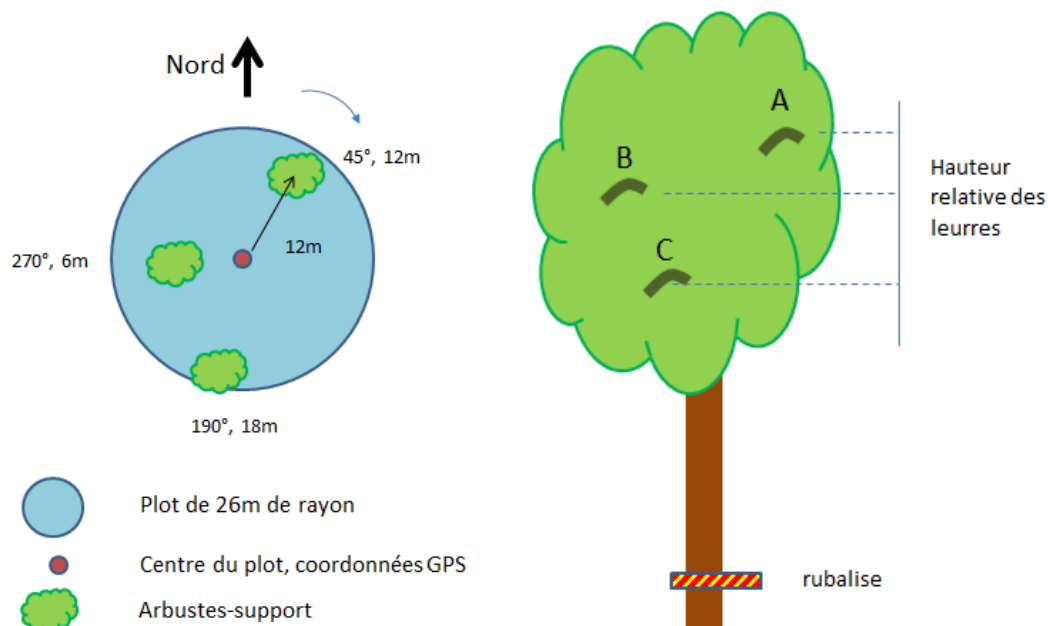


Figure 12 Protocole des relevés du taux de prédation des chenilles.

Tableau 3 Récapitulatif des variables écologiques à disposition et échelle des relevés terrains.

Echelle d'analyse	Données écologiques
Bois	IBP (campagne MUESLI) Abondance/Richesse en oiseaux Indice d'abrutissement par le chevreuil Indice de prédation des défoliateurs
Plots (1ha)	IBP (campagne nationale de 2011)
Plots (rayon 26m)	Indice de prédation des défoliateurs
Plots (rayon 10m)	Richesse botanique

III. Données LiDAR (et hyperspectrales)

Le levé LiDAR aéroporté programmé pour le projet MUESLI a été effectué sur une zone de 260 km² couvrant 28 des 30 bois où des relevés terrains ont été réalisés (Figure 13). A noter que ce plan de vol a été complété par une acquisition d'images hyperspectrales à une résolution de 1m dont on a extrait un indice de végétation (Normalized Difference Vegetation Index, NDVI) que nous utiliserons pour compléter nos modèles statistiques (section II partie 4). Les conditions d'acquisition des levés LiDAR sont visibles en figure 14.

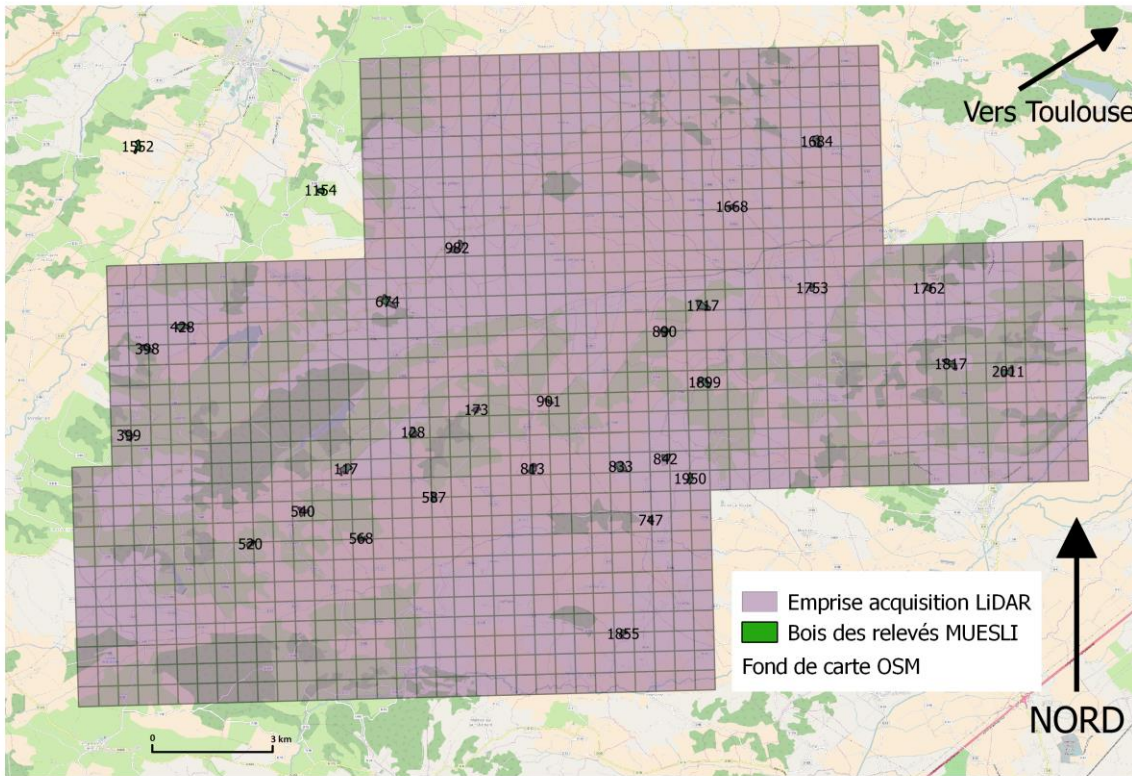


Figure 13 Zone des levés LiDAR et situation des bois relevés.

Date d'acquisition :	15/05/2016	16/05/2016
Laser :	Riegl LMS Q680i	Riegl LMS Q680i
Conditions météorologiques :	nuageux, vent modéré	nuageux, vent modéré
Heure de vol : GPS standard time	14h48 à 16h16	14h03 à 16h46

Paramètres	Valeur nominale
Hauteur de vol	600 m / sol
Vitesse de vol	140 Nœuds
Angle de scan	60°
Précision altimétrique (écart-type)	10 cm
Précision planimétrique (écart-type)	30 cm
Densité native de points	4 points / m ²

Figure 14 Caractéristiques des levés LiDAR issues du cahier des charges.

Les données fournies par le prestataire chargé des relevés comprennent un semis de points déjà classés (à l'aide du logiciel TerraSolid), un Modèle Numérique de Terrain (MNT), un Modèle Numérique de Surface (MNS) et un Modèle Numérique de Hauteur (MNH), tous les trois à 1m de résolution spatiale.

IV. Sortie terrain

Afin de mieux se rendre compte des forêts étudiées, de leurs caractéristiques et de la cohérence entre leur physionomie et les relevés LiDAR obtenus, nous avons effectué une sortie sur le terrain en milieu de stage.

Cette journée sur le terrain s'est déroulée en compagnie d'une personne ayant participé aux relevés pour le projet MUESLI. Lors de cette sortie aucun protocole spécifique n'a été utilisé pour recueillir des informations mais nous avons fait le choix de nous rendre dans des bois, qui, d'après les valeurs de certaines métriques LiDAR calculées (voir partie 4. section I), présentaient des caractéristiques marquées (valeurs extrêmes dans la distribution des valeurs de la métrique concernée).

4. MODELISATION DE VARIABLES ECOLOGIQUES A PARTIR DE VARIABLES LIDAR

A noter que contrairement à la structure « classique » d'articles scientifiques, cette partie présente à la fois la démarche méthodologique mise en œuvre au cours du stage ainsi que les résultats obtenus. Ce choix a été fait car la partie suivante concernant l'extraction des cimes et couronnes des arbres se démarque de cette première partie sur la modélisation statistique. Ainsi, les deux volets (dont le temps passé pour le second au cours du stage est nettement moins important) commencent par une explication de la démarche, directement suivis de la présentation des résultats de l'étude.

La démarche de modélisation effectuée suit différentes étapes, partant du contrôle de la donnée à traiter à la spatialisation des variables d'intérêts grâce à la construction de modèles statistiques à l'aide de variables LiDAR expliquant potentiellement les indicateurs de biodiversité (Figure 15).

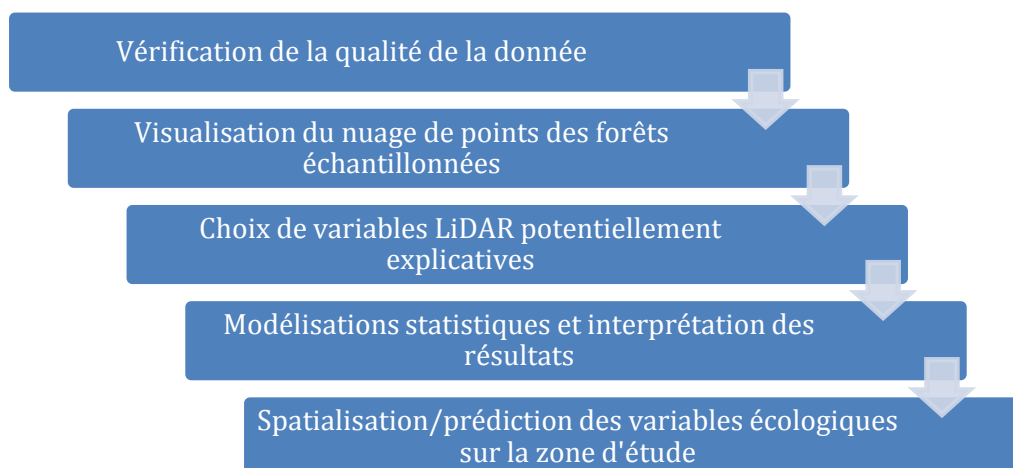


Figure 15 Schéma des étapes de la modélisation statistique, de l'extraction des données LiDAR à la spatialisation des variables écologiques.

I. Démarche d'extraction des variables LiDAR

Le processus d'extraction des variables LiDAR s'effectue en 8 étapes (Figure 16) dont le déroulement permet d'obtenir les variables présentées dans la partie 2 section III afin de pouvoir les injecter dans les analyses statistiques. Nous pouvons voir que, dans un premier temps, l'ensemble des variables LiDAR usuelles sont extraites, puis que d'autres métriques sont calculées. Nous adoptons donc volontairement ici une approche exploratoire c'est-à-dire que nous ne nous restreignons pas à un petit nombre de variables LiDAR qui auraient préalablement été

sélectionnées en raison d'hypothèses écologiques. Nous traitons ainsi un sujet nouveau sans a priori. Cependant, nous calculons aussi d'autres variables LiDAR qui, elles, pourraient expliquer en particulier certains indicateurs ce qui nous permet de complexifier nos modèles.

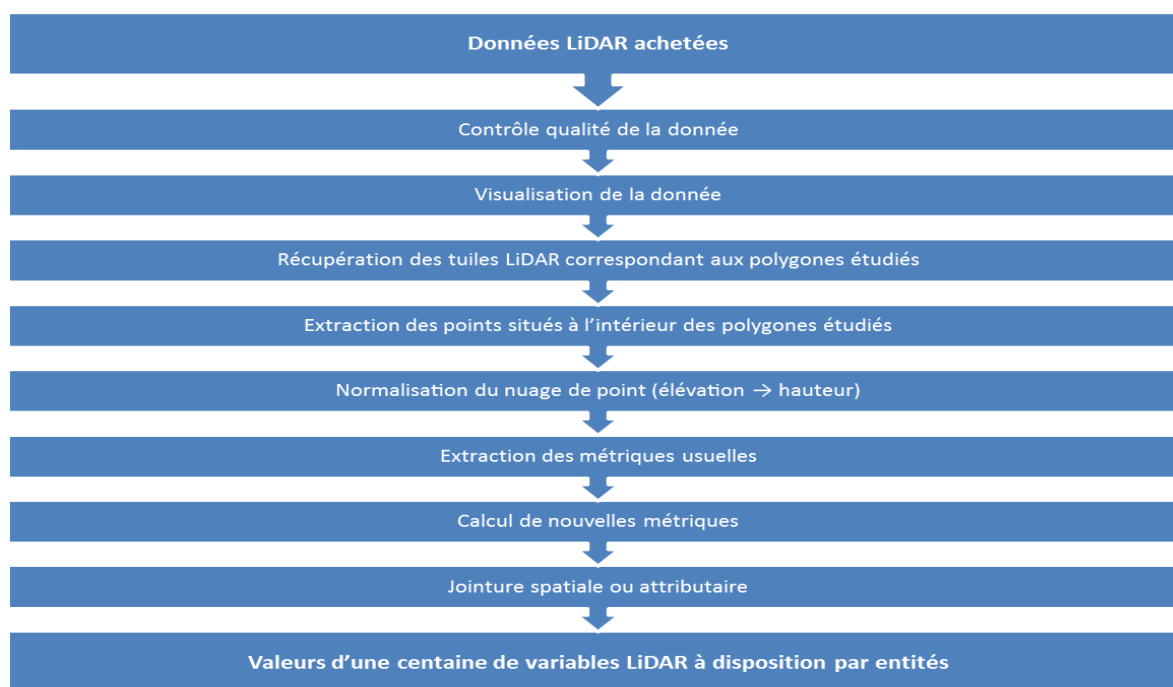


Figure 16 Processus d'extraction des variables LiDAR.

Un contrôle qualité a été effectué en début d'utilisation des données afin de vérifier qu'elles correspondaient bien aux spécifications d'acquisition détaillées dans le rapport technique des levés. Il a été réalisé à l'aide du logiciel LAsTools en suivant le tutoriel¹¹ dédié. Il débute par une visualisation des données afin de vérifier qu'elles ont du sens puis la création d'un index afin d'accélérer les requêtes. La suite du contrôle qualité permet de vérifier le recouvrement interbandes, la précision d'élévation des points LiDAR à partir de points de contrôle, et d'obtenir des informations sur le nuage de points tels la date de création du fichier, le nombre de retours, l'emprise des données, la densité de points et de potentiels messages d'erreur. Il en résulte que les données sont conformes au cahier des charges en termes de précision et de recouvrement. De plus, la densité de points semble un peu plus importante qu'annoncée dans les paramètres de vol. En effet, la densité de points effective est d'environ 10 pts/m² contre 4 pts/m² spécifiés.

Nous avons ensuite cherché à visualiser la donnée, action importante dans ce travail de recherche. Effectivement, elle permet d'une part d'avoir un meilleur aperçu de la donnée et par conséquent de nous fournir des idées sur les différentes informations que nous pouvons extraire et

¹¹ <https://rapidlasso.com/2013/04/20/tutorial-quality-checking/>

leur utilité dans le cadre de nos variables écologiques et d'autre part de valider visuellement les informations extraites. Comme nous l'avons dit précédemment, le logiciel FugroViewer est principalement utilisé pour la visualisation des données LiDAR étudiées dans le cadre de ce stage.

Nous passons ensuite à l'extraction à proprement dit. Les logiciels utilisés pour extraire des variables LiDAR s'exécutent principalement en lignes de commandes. Des scripts, au format batch (.bat), ont donc été créés pour lancer plusieurs commandes à la suite et ainsi effectuer automatiquement l'extraction des variables LiDAR. Des exemples de ces scripts sont fournis en annexe 4.

Les deux logiciels évoqués ont été utilisés conjointement afin d'optimiser les calculs et les fonctions de chacun. Plus concrètement, les calculs se déroulent dans un dossier contenant les tuiles LiDAR à utiliser ainsi que le shapefile définissant nos zones d'étude (i.e. plots ou bois). Pour simplifier les calculs, les tuiles LiDAR utiles (i.e. celles intersectant les polygones en entrée) sont présélectionnées manuellement. A noter qu'il a été nécessaire de retracer les polygones correspondant aux bois. Certaines entités contenues dans la couche préalablement fournie comportant des zones hors forêt ou des lignes électriques coupant les bois (Figure 17), ce qui faussait les résultats obtenus. En effet, les valeurs des variables LiDAR obtenues étaient biaisées car ne correspondaient pas totalement à des zones forestières alors que les mesures effectuées sur le terrain ne tenaient compte que de la végétation.

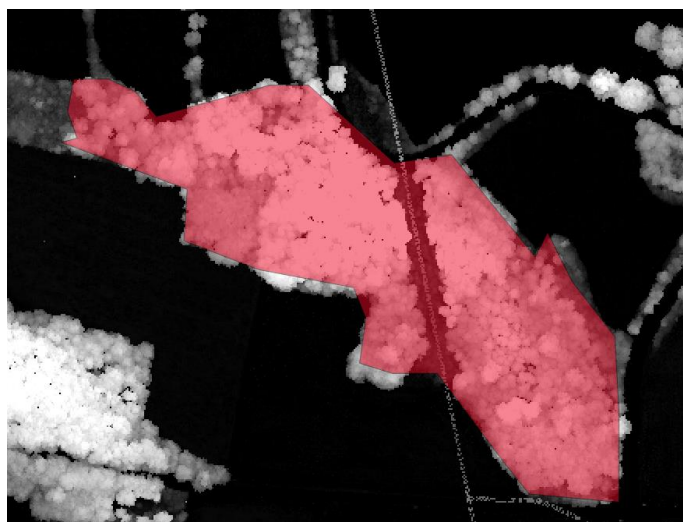


Figure 17 Exemple d'un bois coupé par une ligne électrique et ayant dû être retracé.

La première étape de l'extraction vise elle aussi à simplifier les calculs. A savoir qu'elle consiste à découper les tuiles utiles pour ne conserver que les données LiDAR à l'intérieur des polygones. Cela est réalisé avec FUSION ; LAsTools le permettant aussi mais n'est utilisable en version

gratuite que sur une petite quantité de données sinon des informations, notamment les valeurs d'intensité, sont mises à zéro. C'est d'ailleurs à cause de cette contrainte logicielle que nous traiterons d'abord les variables liées à l'intensité puis celle liées à la hauteur. En effet, l'étape de normalisation du nuage de points nécessaire à l'extraction des variables de hauteur n'est possible qu'avec LAStools et peut donc induire une perte des valeurs d'intensité.

Nous normalisons donc ensuite le nuage de points. Cela permet d'obtenir des valeurs de hauteur pour chaque point LiDAR à la place des valeurs d'altitude, par soustraction des valeurs d'altitude du terrain (issue du MNT LiDAR).

Puis un index spatial des tuiles normalisées est créé permettant d'accélérer les calculs par la suite. Et enfin, les métriques forestières classiques sont extraites en prenant en entrée les données normalisées créées précédemment. Les résultats sont stockés dans des fichiers tableurs (au format csv) à l'intérieur d'un dossier résultat créé à cet effet. Certaines variables issues des deux logiciels sont communes. Nous avons alors comparé leurs valeurs qui s'avèrent identiques d'un logiciel à l'autre.

D'autres métriques sont aussi calculées. Pour cela, nous filtrons tout d'abord le nuage de points afin de ne conserver que les premiers échos puis nous extrayons la variance de l'intensité car ce sont ces échos qui fournissent l'intensité la plus fiable (Morsdorf et al. 2010). Ensuite, nous calculons la proportion de premiers échos de faible intensité par rapport à l'ensemble des premiers échos. Puis, nous calculons la densité de canopée à différentes hauteurs (2m pour le seuil arbustif et 7m pour le seuil arboré) ainsi qu'un indice de complexité verticale de la canopée (Vertical Complexity Indice, VCI en anglais) en prenant en compte quatre strates ($h < 2m$, $2m < h < 7m$, $7m < h < 20m$ et $h > 20m$). Il est défini comme suit : $\frac{\sum_{i=1}^{Ns} Ni \cdot \ln(Ni)}{\ln(Ns)}$ où Ns correspond au nombre de strates verticales considérées (ici 4 donc) et Ni aux proportions de retours pour chaque strate de hauteur considérée. Comme nous l'avons évoqué, ces choix de variables sont faits en tenant compte des indicateurs de biodiversité à notre disposition. Wing et al. (2015) montrent que les arbres morts renvoient une intensité LiDAR plus faible. Nous cherchons donc, en calculant les échos de faible intensité, à expliquer l'IBP car certains facteurs permettant de le calculer tiennent compte du bois mort. Le VCI peut notamment être intéressant pour tenir compte du nombre de strates verticales dans le calcul de l'IBP ou bien encore pour voir si l'abrutissement lié au chevreuil est lié à l'encombrement d'une forêt.

Enfin, il est nécessaire de regrouper ces variables et de les rattacher à leurs polygones d'origine, afin d'effectuer les analyses statistiques par la suite. Cette dernière étape se fait sur un logiciel SIG

(ici QGIS) en réalisant des jointures (attributaire pour celles provenant de FUSION à l'aide d'un identifiant stocké dans le tableur et spatiale pour les métriques issues de LAStools à l'aide des coordonnées du centroïde de l'entité). Nous obtenons ainsi une centaine de variables LiDAR : différentes statistiques descriptives sur la hauteur et l'intensité des objets avec les deux logiciels, une vingtaine de pourcentages attestant du nombre de retour au-dessus d'un seuil, propres à FUSION, les densité et couverture de canopée propres à LAStools et trois autres variables calculées à partir de FUSION (VCI, proportion de premiers échos de faible intensité par rapport à l'ensemble des premiers échos, variance de l'intensité des premiers retours).

II. Modélisation statistique

1) Exploration des données

Une fois les variables LiDAR extraites il convient d'établir des relations empiriques entre celles-ci et les variables écologiques à disposition pour mesurer ou non un effet de la structure du milieu sur les indicateurs de biodiversité mesurés. Pour cela, nous avons choisi deux méthodes statistiques permettant d'expliquer et prédire une variable d'intérêt à partir de variables explicatives. Il s'agit des modèles linéaires généralisés (Generalized Linear Model, GLM) qui feront l'objet de la deuxième section et de la régression par les moindres carrés partiels (Partial Least Squares, PLS). Les résultats de la PLS n'étant pas significatifs, nous avons choisi de ne pas faire figurer cette méthode dans le corps de ce rapport. L'intérêt et les résultats obtenus avec cette méthode sont toutefois fournis en annexe 7.

Au préalable, il est nécessaire d'explorer les différentes variables à disposition afin de garantir la bonne utilisation des modèles statistiques. Pour ce faire, nous avons adopté le protocole proposé par Zuur et al. (2010) dédié aux applications écologiques. Ce protocole émet des recommandations importantes afin d'éviter de tirer de fausses conclusions des analyses statistiques réalisées. Il se déroule en huit étapes vérifiant chacune une propriété de la distribution des variables à expliquer, des variables explicatives ou de leurs interactions (Figure 18).

La première étape consiste à supprimer les données aberrantes (*outliers* en anglais). Il s'agit d'observations dont les valeurs sont extrêmes par rapport aux autres. Une représentation des valeurs prises par la variable à expliquer sous forme de boîte à moustache faisant figurer notamment la médiane et les 1^{er} et 3^{ème} quartiles permet d'identifier ces valeurs. Les valeurs extrêmes, en dehors de la boîte et des moustaches, doivent alors être étudiées afin de voir s'il s'agit effectivement de données aberrantes (erronées) qu'il convient d'écarter.

Ensuite, il faut s'assurer que la variance de la variable à expliquer est bien homogène. Dans le cas de modèles de type régression, cela se vérifie par une variation des résidus du modèle calculé similaire pour chaque valeur ajustée. Dans le cas contraire il peut être intéressant de transformer la variable réponse afin que la variance ne soit plus hétérogène.

Il faut aussi connaître la distribution de la variable réponse avant de choisir la modélisation statistique à mettre en place. En effet, certaines méthodes supposent une distribution normale de cette variable réponse (ou des résidus) comme c'est le cas de la régression linéaire.

Une analyse plus en détail des données doit être effectuée lorsque l'exploration des données révèle de nombreuses valeurs nulles. En effet, de telles distributions peuvent entraîner des résultats biaisés et compliquent l'interprétation des analyses statistiques.

La cinquième étape revient à évaluer la corrélation entre les variables explicatives (i.e. pour nous, les variables LiDAR). Même si théoriquement nous pourrions utiliser toutes les variables explicatives à notre disposition pour créer un modèle statistique, si certaines variables sont corrélées entre elles les résultats seront faussés. L'utilisation de variables colinéaires peut rendre difficile l'évaluation des effets indépendants de chacune des variables puisque l'influence de l'une d'entre elles dépend aussi de celles des autres. Leur effet pourrait aussi ne pas être détecté (Dormann et al. 2013). Il est donc plus juste d'utiliser des variables explicatives peu corrélées pour créer les modèles de régression.

Ensuite, il est aussi important de visualiser la relation entre la variable réponse et chaque variable explicative. Pour cela il convient de représenter graphiquement la variable à expliquer en fonction de chaque variable à disposition (*scatter plot*). Nous pouvons alors remarquer si des valeurs se distinguent au travers du nuage de points des observations et qui nécessiteraient une analyse complémentaire. Nous pouvons également observer si une relation linéaire apparaît ou s'il convient d'envisager une modélisation non linéaire.

La septième étape du protocole consiste à regarder s'il y a des interactions entre les variables. Complémentairement à l'étape 5, nous ne regardons plus la corrélation entre deux variables mais entre un groupe, de trois ou plus. Deux variables peuvent alors avoir une influence simultanée sur une troisième. Cela implique des difficultés d'interprétation car l'influence d'une variable dépend de l'influence de son interaction avec d'autres. Concernant ce point, nous n'avons pas tenu compte de ces interactions dans ce stage. Le travail effectué constitue donc un premier niveau d'analyses.

La dernière étape du processus d'exploration des données concerne l'indépendance des observations de la variable à expliquer. Il n'est pas rare d'avoir des échantillons dépendants avec des données géographiques en raison des relations spatiales existantes. Dans le cadre de notre étude par exemple, il peut s'agir d'observations effectuées à l'échelle d'un plot mais dont plusieurs sont issus d'une même forêt. Les résultats peuvent donc être influencés par cette proximité géographique impliquant potentiellement des structures de végétation dépendantes. Nous verrons par la suite comment passer outre ce problème dans la construction de nos modèles statistiques par GLM.

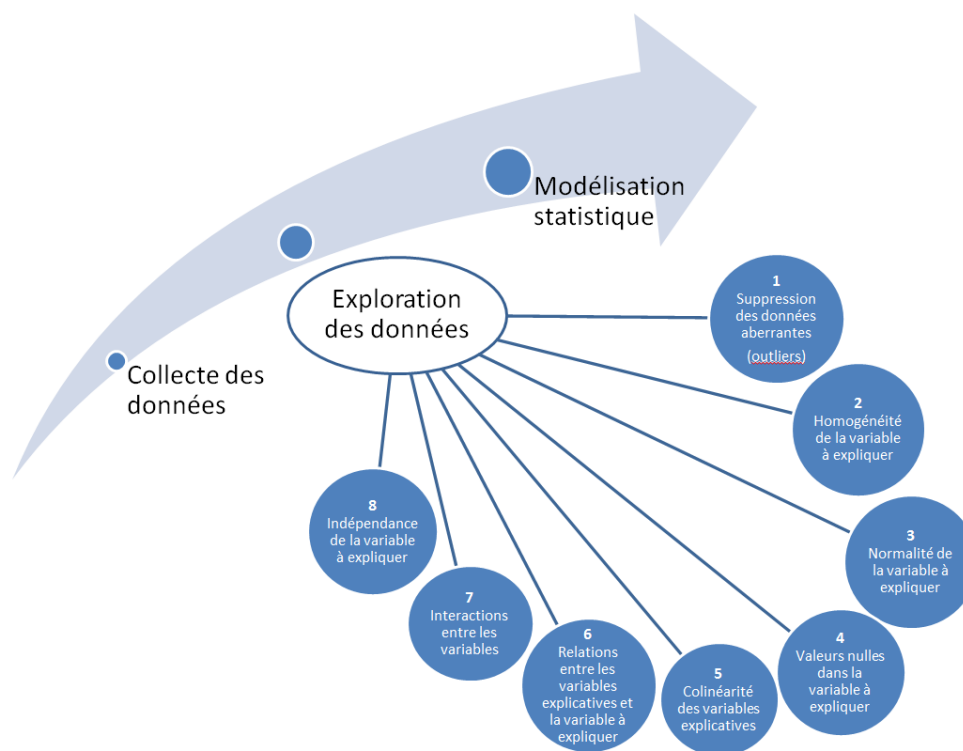


Figure 18 Schéma des étapes d'exploration des données à réaliser avant une modélisation statistique.

Nous avons donc exploré nos données en suivant ces préconisations avant d'effectuer les modélisations statistiques. La réalisation des étapes 1 et 5 a nécessité davantage de traitements, nous les détaillerons donc ici. En effet, nous avons tout d'abord noté la présence d'une hauteur maximale d'arbre anormalement élevée. Il fallait alors vérifier s'il s'agissait d'une valeur aberrantes ou non. Le retour étant situé dans une zone de hauteurs d'arbres homogènes il était certainement dû à la présence d'un oiseau dans les airs. La hauteur a donc été recalculée après avoir supprimé le point en question jugé aberrant. Nous avons ensuite regardé la corrélation entre les variables explicatives LiDAR pour n'en conserver qu'un panel peu corrélées entre elles (seuil fixé à 0,7 avec significativité en suivant les recommandations de Dormann et al. 2013). Il s'avère que sur la

centaine de variables LiDAR à notre disposition, beaucoup étaient fortement corrélées entre elles. Ainsi, nous avons retenu 10 variables peu corrélées (i.e. inférieur au seuil de 0.7 préconisé) à l'échelle des bois, 12 à l'échelle des plots d'1ha, 11 à l'échelle des plots de rayons proches de 26m et 15 à l'échelle des plots de 10m de rayon (Tableau 4).

Tableau 4 Variables LiDAR peu corrélées à chaque échelle d'analyse.

Echelle d'analyse	Variables LiDAR peu corrélées
Bois	Hauteur maximale Hauteur moyenne Coefficient d'aplatissement (Kurtosis) Intensité maximale Nombre de retours supérieurs à 3m Relief de la canopée $\left(\frac{H_{moy} - H_{min}}{H_{max} - H_{min}}\right)$ Inverse de la densité de canopée avec une coupe de hauteur à 2m Couverture de canopée avec une coupe de hauteur > 2m Proportion de 1ers retours de faible intensité par rapport au total 1ers retours Indice de complexité verticale (VCI)
Plots (1ha)	Hauteur maximale Hauteur moyenne Nombre de 5 ^{ème} retour Coefficient d'aplatissement (Kurtosis) Percentile 1% Percentile 99% Intensité maximale Variance de l'intensité Nombre de retours supérieurs à 3m Couverture de canopée avec une coupe de hauteur > 2m Proportion de 1ers retours de faible intensité par rapport au total 1ers retours Indice de complexité verticale (VCI)
Plots (rayon 26m)	Hauteur maximale Hauteur moyenne Coefficient de dissymétrie (Skewness) Coefficient d'aplatissement (Kurtosis) Percentile 1% Percentile 99% Nombre de retours supérieurs à 3m Couverture de canopée avec une coupe de hauteur > 2m Couverture de canopée avec une coupe de hauteur > 7m Proportion de 1ers retours de faible intensité par rapport au total 1ers retours Indice de complexité verticale (VCI)
Plots (rayon 10m)	Hauteur maximale Variance de hauteur Coefficient d'aplatissement (Kurtosis) Percentile 1% Intensité maximale Variance de l'intensité des 1 ^{ers} retours Nombre de 5 ^{ème} retour Pourcentage de 1 ^{er} retour au-dessus de la hauteur moyenne Pourcentage de 1 ^{er} retour au-dessus du mode de hauteur Relief de la canopée $\left(\frac{H_{moy} - H_{min}}{H_{max} - H_{min}}\right)$ Couverture de canopée avec une coupe de hauteur > 2m Couverture de canopée avec une coupe de hauteur > 7m Inverse de la densité de canopée avec une coupe de hauteur > 2m Proportion de 1ers retours de faible intensité par rapport au total 1ers retours Indice de complexité verticale (VCI)

2) Modélisation statistique par GLM

Nous cherchons tout d'abord à expliciter le lien entre les variables LiDAR extraites et les indicateurs écologiques à l'aide de modèles linéaires généralisés basés sur un modèle de régression. Ces modèles se différencient des modèles linéaires classiques. Ils permettent en effet de prendre en considération plusieurs types de distributions de la variable réponse (par exemple loi binomiale ou loi de Poisson) tandis que les modèles linéaires classiques ne sont utilisables que pour une variable à expliquer suivant une loi normale. Les paramètres du modèle sont alors estimés grâce à la méthode du maximum de vraisemblance (ou dite du minimum de déviance) et non plus par la méthode des moindres carrés (Guisan et al. 2002).

Les modèles créés sont résumés dans le tableau 5 (les numéros donnés seront les mêmes pour les résultats). Nous cherchons à ajuster un modèle de la forme :

$g(E(y)) = b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n + \varepsilon$, où (X_1, X_2, \dots, X_n) sont les variables explicatives (issues de la télédétection en particulier, y représente la variable réponse à expliquer (indicateur de biodiversité) et ε les résidus.

Nous ajustons d'abord des modèles qui intègrent seulement des variables LiDAR. Nous regardons ainsi les effets potentiels de la structure de la végétation (notamment verticale) sur la variable réponse. Pour l'IBP, par exemple, nous savons que seule une partie de son calcul est lié à cette structure (facteur lié au peuplement et à la gestion forestière). C'est pourquoi nous testons ensuite des modèles liant seulement cette partie gestion aux variables LiDAR (modèle 4 tableau 5) puis nous utilisons les notes liées au contexte en tant que variables explicatives (modèle 6 tableau 5). Nous complexifions aussi les modèles en ajoutant des variables explicatives extérieures au LiDAR. Le NDVI est injecté afin de caractériser la couverture de végétation et son hétérogénéité. Des facteurs écologiques sont aussi utilisés en tant que variables explicatives. Nous supposons que la présence et la diversité en oiseaux influe sur la prédation des chenilles et nous savons que la richesse en oiseaux augmente avec la surface des bois. Nous voulons ainsi voir si ces variables, combinées au LiDAR, ont une influence significative sur les indicateurs de biodiversité étudiés.

Tableau 5 Modèles GLM calculés.

N°	Modèles créés	
	Variables à expliquer	Variables explicatives
	Données IBP :	
1	IBP total (données MUESLI)	Variables LiDAR
2	IBP total (données MUESLI)	Variables LiDAR, NDVI moyen, NDVI variance
3	IBP total - données MUESLI)	Variables LiDAR, richesse en oiseaux nichant dans des cavités, log(surface bois)

4	IBP partie gestion (données 2011)	Variables LiDAR
5	IBP total (données 2011)	Variables LiDAR
6	IBP total (données 2011)	Variables LiDAR, notes IBP contexte
Données prédation des chenilles :		
7	Prédation chenilles (échelle du bois)	Variables LiDAR
8	Prédation chenilles (échelle du bois)	Variables LiDAR, abondance en oiseaux
9	Prédation chenilles (échelle du bois)	Variables LiDAR, NDVI moyen, NDVI variance
10	Prédation chenilles (échelle du bois)	Variables LiDAR, NDVI moyen, NDVI variance, abondance en oiseaux insectivores
11	Prédation chenilles (échelle du bois)	Variables LiDAR, NDVI moyen, NDVI variance, richesse en oiseaux insectivores
12	Prédation chenilles (échelle du bois)	Variables LiDAR, NDVI moyen, NDVI variance, abondance en oiseaux s'alimentant dans la canopée
13	Prédation chenilles (échelle du bois)	Variables LiDAR, NDVI moyen, NDVI variance, richesse en oiseaux s'alimentant dans la canopée
14	Prédation chenilles (échelle des plots)	Variables LiDAR
Données indice de consommation par le chevreuil :		
15	Indice de consommation par le chevreuil (échelle du bois)	Variables LiDAR
Données richesse de la strate herbacée :		
16	Richesse strate herbacée (échelle des plots)	Variables LiDAR
Données richesse en oiseaux :		
17	Richesse totale en oiseaux	Variables LiDAR, log(surface bois)
18	Richesse en oiseaux forestiers	
19	Richesse en oiseaux nichant dans des cavités	
20	Richesse en oiseaux nichant au sol	
21	Richesse en oiseaux nichant dans la canopée	
22	Richesse en oiseaux nichant dans la strate basse	
23	Richesse en oiseaux s'alimentant dans la canopée	
24	Richesse en oiseaux s'alimentant au sol	
25	Richesse en oiseaux s'alimentant dans les strates basses	
26	Richesse en oiseaux insectivores	

Pour pouvoir appliquer ce type de modélisation statistique, deux conditions doivent être en particulier vérifiées : l'homogénéité de la variance et la répartition aléatoire des résidus du modèle de régression (cf. infra). Une fois ces conditions validées, une famille doit être choisie correspondant à la nature de la variable à expliquer (e.g. gaussienne, poisson, binomiale). Pour cela, nous effectuons un test de Shapiro afin de tester l'hypothèse nulle selon laquelle la variable à expliquer est normalement distribuée. Si la p-value de ce test est significative (seuil alpha fixé à 0.05) alors nous ne pouvons pas utiliser une loi normale. Il pourrait alors s'agir d'une loi de Poisson. Nous réalisons alors un test de sur-dispersion pour le vérifier. Il est cependant important de noter que, suivant la nature de la variable elle devrait théoriquement suivre une certaine loi mais, en pratique, cela n'est pas toujours vérifié dans les données à la suite de ces tests.

A la suite de différents tests et après discussion avec un statisticien du laboratoire Dynafor, nous pouvons dresser le tableau 6 concernant le choix de la distribution adaptée à la variable à expliquer.

Tableau 6 Distribution des variables réponses.

Variable écologique	Type	Famille
IBP totale	Note (comptage)	Gaussienne
IBP partie gestion	Note (comptage)	Gaussienne (après avoir rendu sa distribution symétrique)
Prédation des chenilles	Taux	Binomiale
Abroustissement par le chevreuil	Taux	Gaussienne
Richesse spécifique en oiseaux	Comptage	Poisson
Richesse botanique	Comptage	Poisson

Ensuite, nous mettons en place sous **R** l'analyse GLM entre la variable à expliquer (*i.e.* une des variables écologiques) et les variables explicatives (*i.e.* les variables Lidar). Le meilleur modèle est alors choisi avec la procédure « stepAIC » permettant de calculer le critère d'information d'Akaike (Akaike Information Criterion, AIC en anglais) pour l'ensemble des combinaisons de variables et de conserver le plus parcimonieux. Cette procédure se base sur la formule $AIC = -2 \cdot \ln(L) + 2 \cdot k$ avec L la vraisemblance maximisée et k le nombre de paramètres du modèle (*i.e.* le nombre de variables). A noter qu'elle n'est valable que lorsque $n/k < 40$ avec n correspondant au nombre d'observations. L'indice AIC le plus bas est alors choisi en considérant qu'un modèle est meilleur qu'un autre si la différence d'AIC est supérieure à deux unités ($\Delta AIC \geq 2$) sinon ils sont considérés comme équivalents. Les variables permettant d'obtenir le modèle avec le plus faible AIC (soit le modèle le plus parcimonieux) sont alors utilisées pour construire le modèle final. Nous regardons alors la significativité des variables explicatives choisies à l'aide de la p -value et la déviance du modèle (pseudo- R^2) est calculée. Cette dernière fournit une indication sur la performance explicative du modèle. A noter que nous avons pris en compte la huitième et dernière étape d'exploration des données concernant l'indépendance des observations. Effectivement, nous pouvons considérer que les plots issus d'une même forêt sont potentiellement dépendants car les forêts sont pour la plupart relativement homogènes. Nous avons donc pris en considération cette proximité géographique dans la construction des modèles à l'échelle des plots en utilisant aussi des GLMM (Generalized Linear Mixed Model). Nous ajoutons pour chaque plot le bois d'origine auquel il appartient en introduisant un effet aléatoire dans le modèle sur les bois. Nous utilisons des observations indépendantes et tenons compte de la dépendance spatiale possible des observations.

III. Prédiction et spatialisation des variables écologiques

Les modèles construits précédemment sont surtout explicatifs mais peuvent servir aussi à prédire et spatialiser. Il suffit de calculer les variables LiDAR sur de nouvelles zones puis d'appliquer les modèles. Pour cela nous avons tout d'abord extrait les zones de végétation à partir de la BD TOPO® puis nous avons conservé seulement les entités d'une superficie suffisante après fusion

des tuiles adjacentes. Le but étant d'obtenir des entités se rapprochant de celles correspondant aux observations afin que les modèles statistiques construits soient transposables. Une fois ces entités obtenues nous pouvons extraire des données LiDAR la valeur de chaque variable utile au modèle. Concernant les modèles à l'échelle des plots, les variables LiDAR ont été calculées dans une grille dont la surface de chaque unité se rapproche de celle des plots. Nous prédisons alors des valeurs spatialisées pour quelques indicateurs de biodiversité sur des zones forestières de l'emprise des levés LiDAR. L'intérêt de la démarche n'est pas ici d'obtenir une valeur exacte de l'indicateur en chaque lieu mais plutôt d'obtenir une carte à large échelle, en valeur relative. Cela permet de distinguer des bois dont les valeurs de l'indicateur sont faibles ou élevées ce qui pourrait guider la construction d'un nouvel échantillonnage terrain.

IV. Résultats

Concernant la modélisation statistique par GLM, nous présenterons en détail les résultats d'un seul modèle en prenant comme exemple celui relatif à la prédation des chenilles à l'échelle du bois (modèle n°7). Nous synthétiserons ensuite les résultats de tous les modèles dans un tableau en fin de section. Le détail de chacun de ceux-ci peut être trouvé en annexe 5.

Selon la démarche décrite précédemment, nous commençons par tracer le diagramme de boîte à moustache de l'indice prédation des défoliateurs, notre variable réponse d'intérêt (Figure 19). Nous remarquons qu'il n'y a pas de valeurs extrêmes et que la distribution n'est pas gaussienne.

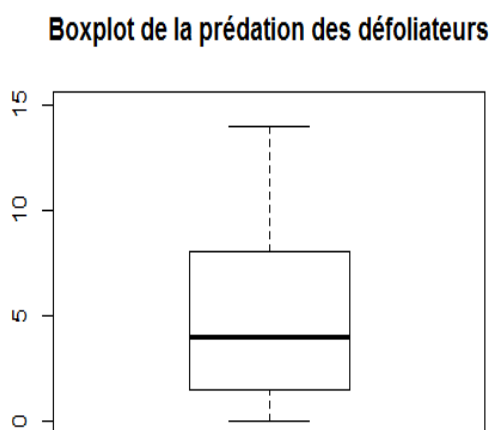


Figure 19 Boîte à moustache de la prédation des défoliateurs à l'échelle des bois.

Après, nous cherchons quelles variables LiDAR sont peu corrélées (< 0.7). Pour cela, nous effectuons un test de corrélation de Pearson et représentons le corrélogramme associé. L'exemple

figure 20 représente les variables non corrélées obtenues à l'échelle du bois. Nous pouvons voir que la corrélation est bien inférieure à 0.7 en valeur absolue.

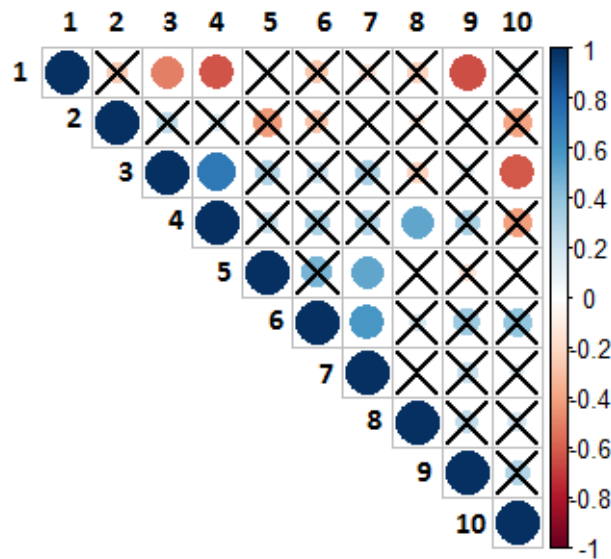
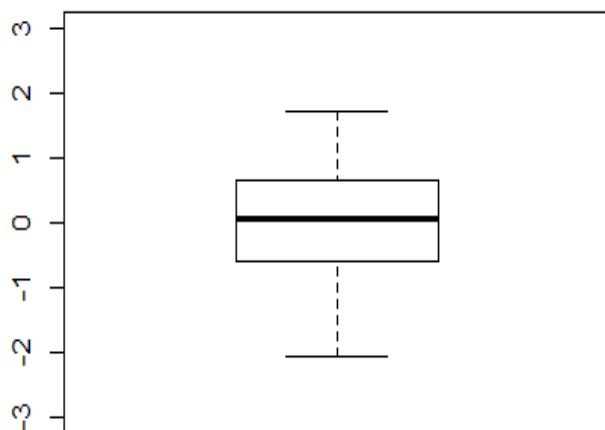


Figure 20 Corrélogramme des variables peu corrélées à l'échelle des bois (croix noire : valeur non significative, taille des cercles proportionnelle aux coefficients de corrélation).

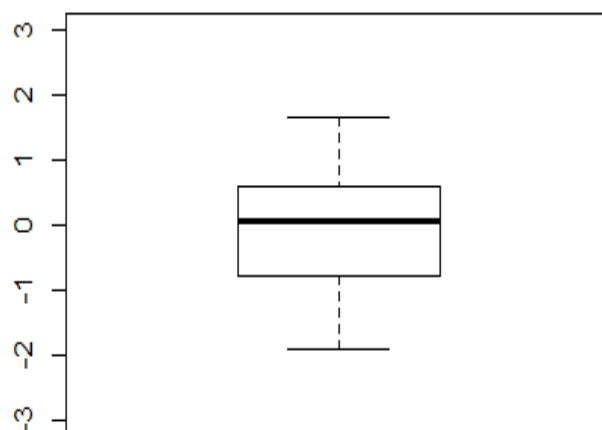
- 1 : Inverse de la densité de canopée avec une coupe de hauteur à 2m
- 2 : Proportion de 1ers retours de faible intensité par rapport au total de 1ers retours
- 3 : Hauteur maximale
- 4 : Hauteur moyenne
- 5 : Intensité maximale
- 6 : Coefficient d'aplatissement (Kurtosis)
- 7 : Nombre de retours supérieurs à 3m
- 8 : Relief de la canopée $\frac{H_{moy} - H_{min}}{H_{max} - H_{min}}$
- 9 : Couverture de canopée avec une coupe de hauteur à 2m
- 10 : Indice de complexité verticale (VCI)

Nous visualisons ensuite les boîtes à moustache de chacune des variables LiDAR retenues (Figure 21) ainsi que la distribution de la variable réponse en fonction de ces données (*scatter plot*). Nous pouvons remarquer sur les graphiques suivants des valeurs potentiellement aberrantes. Cependant, ces valeurs ont été vérifiées individuellement et il se trouve qu'il s'agit de données justes, correspondant à des valeurs extrêmes caractérisant des bois différents des autres par leur structure. Les diagrammes de dispersion quant à eux (Figure 22) ne laissent pas entrevoir des corrélations importantes entre une variable LiDAR et le taux de prédation des défoliateurs.

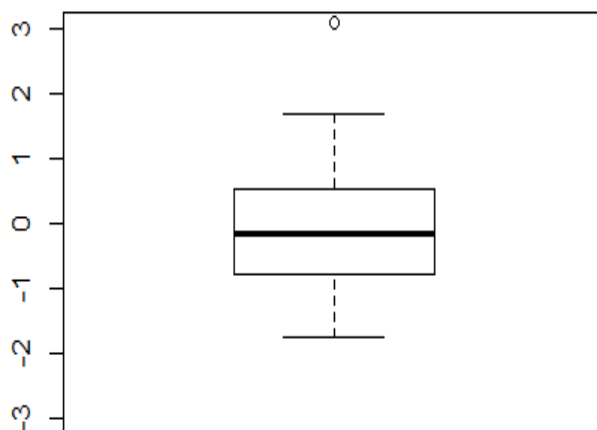
Boxplot de la hauteur maximale



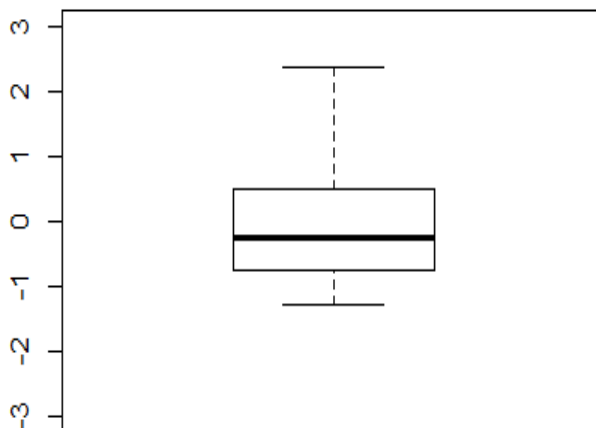
Boxplot de la hauteur moyenne



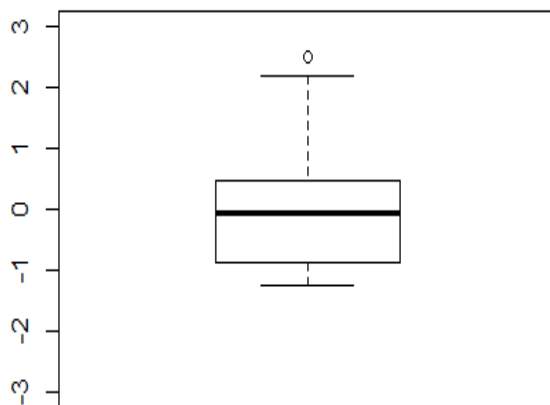
Boxplot du Kurtosis



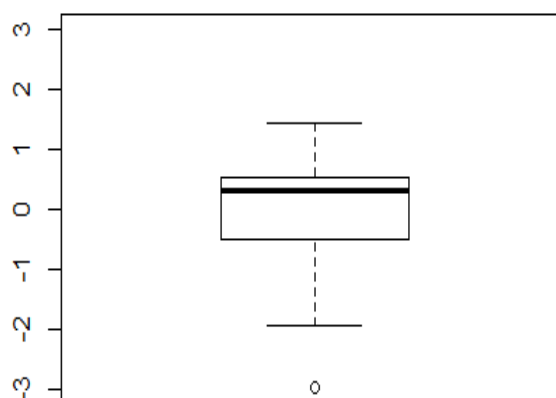
Boxplot de l'intensité maximale



Boxplot de l inverse de la densité de canopée



Boxplot de la couverture de canopée



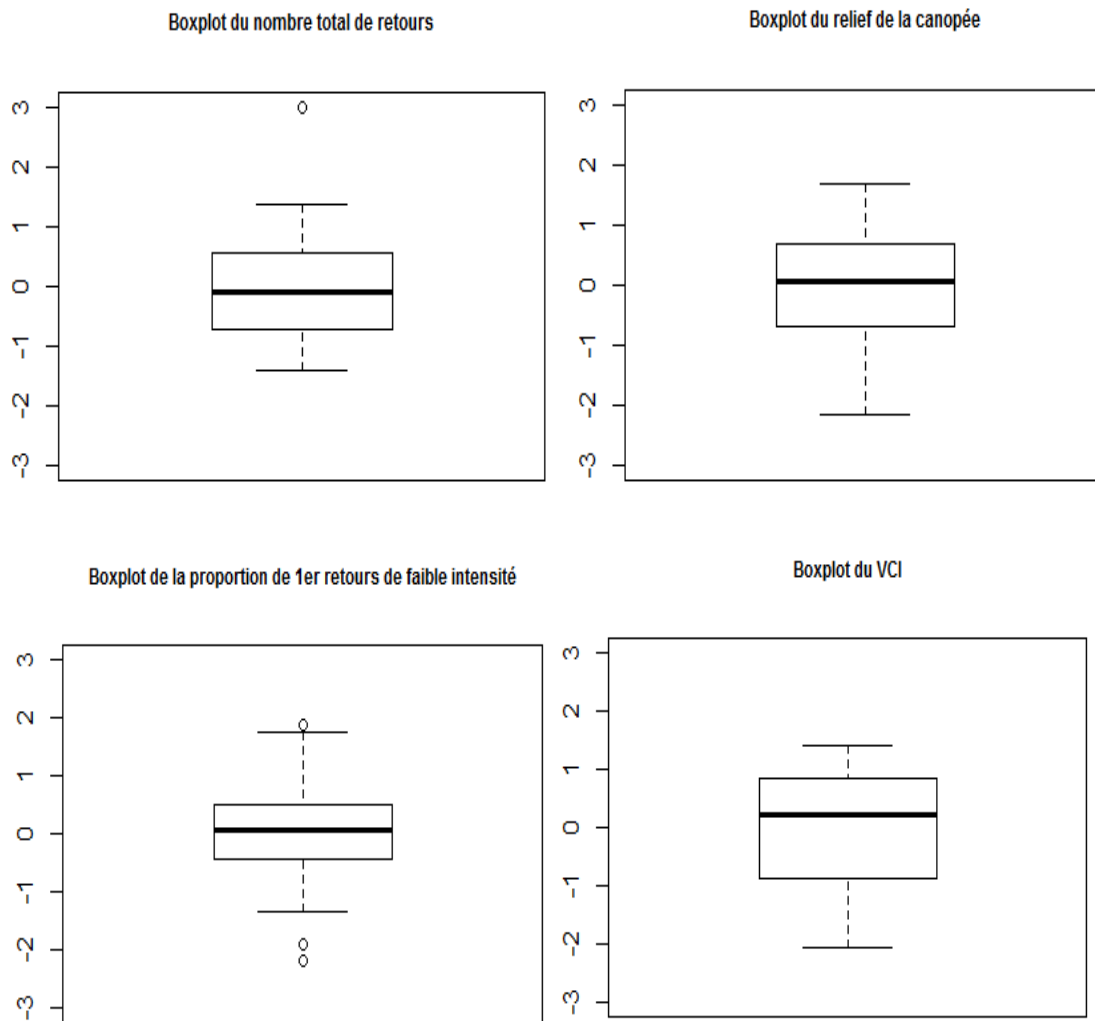
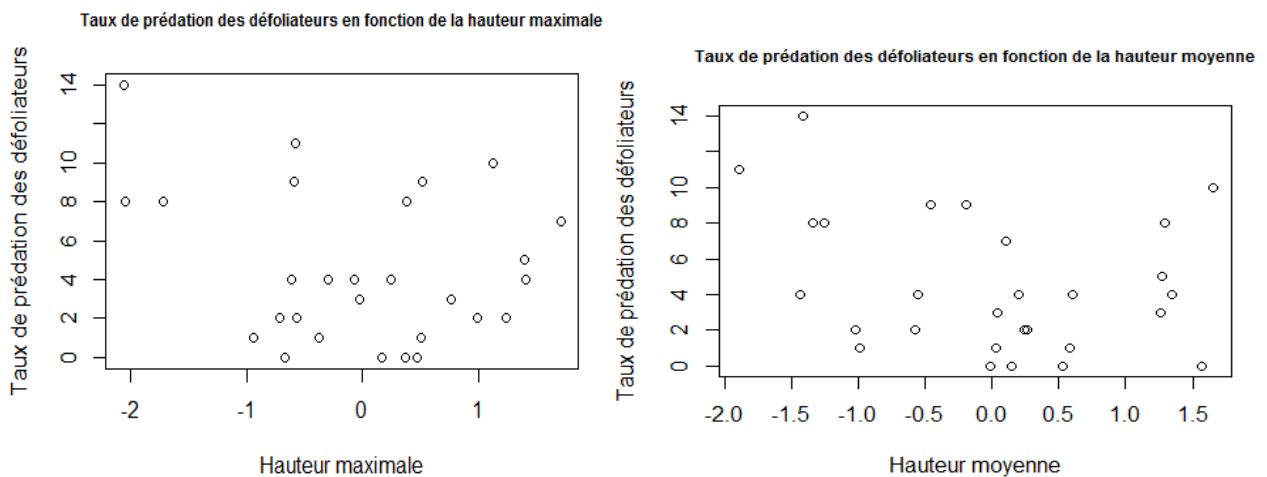
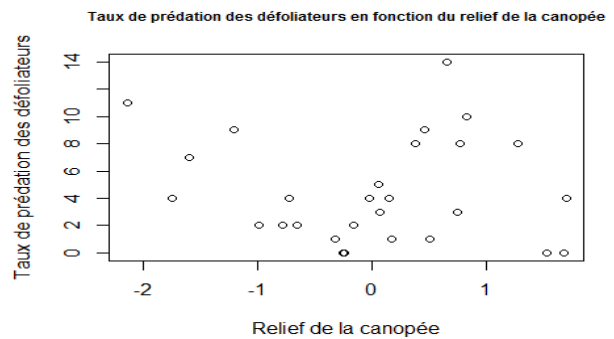
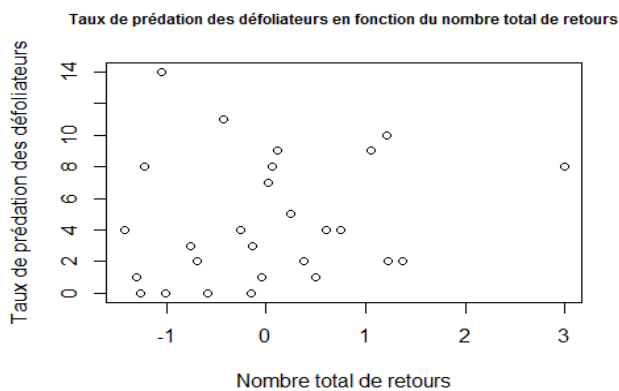
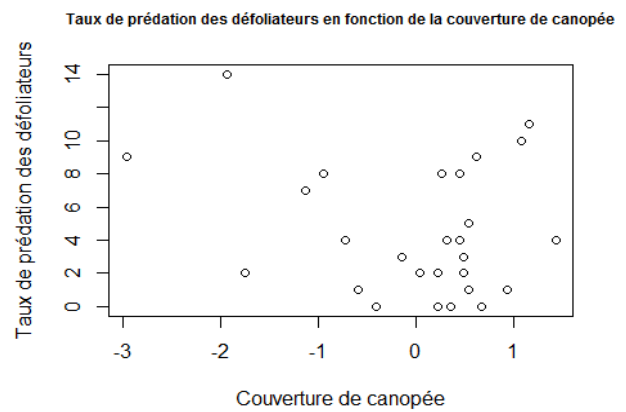
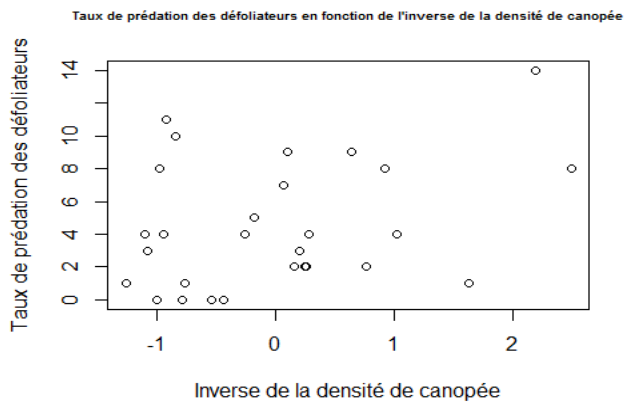
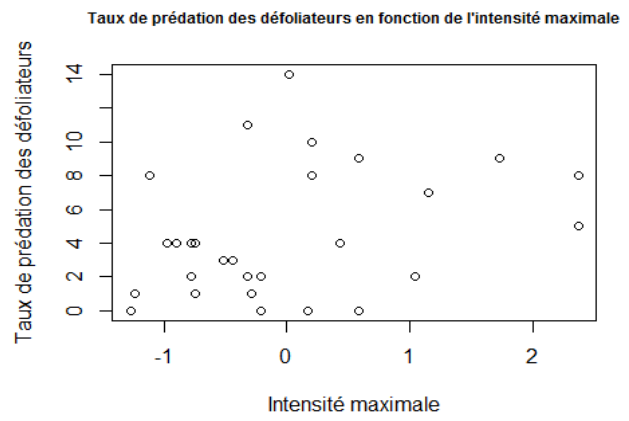
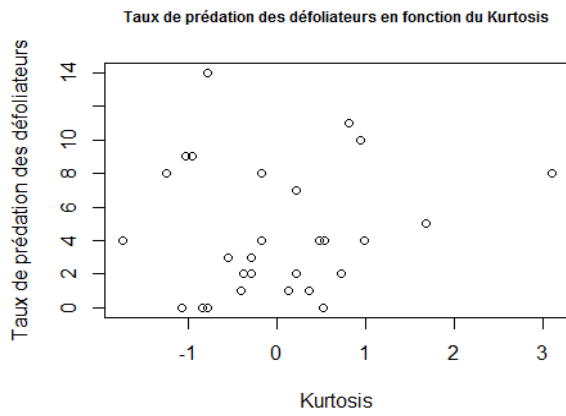


Figure 21 Boîtes à moustaches des différentes variables LiDAR, centrées et réduites, peu corrélées à l'échelle des bois.





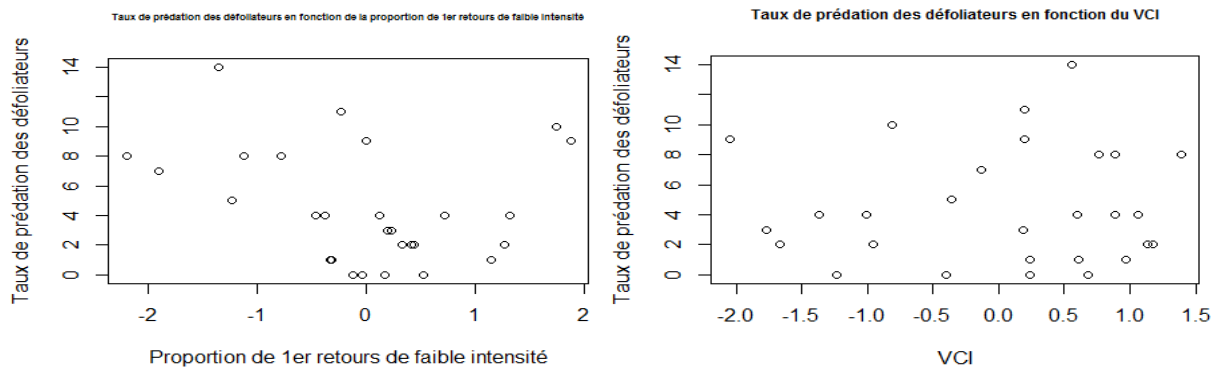


Figure 22 Diagrammes de dispersion du taux de prédation des chenilles en fonction des variables LiDAR peu corrélées à l'échelle des plots.

Enfin, nous sélectionnons le modèle GLM le plus parcimonieux et nous analysons l'effet des variables et la déviance du modèle (exemple ci-dessous de script R pour construire le modèle).

```
## Chenilles à l'échelle du bois (sans abondance en oiseaux)
# Construction du modèle GLM où l'on cherche à expliquer La variable Taux de
# prédation en défoliateurs en fonction des variables LiDAR. La distribution de
# La variable réponse est supposée suivre une loi binomiale.

mod_chenilles =
glm(cbind(dataChenilleUseful$pred,dataChenilleUseful$tot_chenilles-
dataChenilleUseful$pred) ~ cut2_cov + F_Elev.maximum + F_Elev.mean +
F_Elev.kurtosis + F_Int.maximum + cut2_dns_gap + F_Total.all.returns +
F_Canopy.relief.ratio + myTableScale$propFirst + myTableScale$VCI, family =
"binomial", data = dataMetricLidScale)

# Sélection du modèle le plus parcimonieux à partir du modèle intégrant toutes
# Les variables non corrélées

stepAIC(mod_chenilles)

## Step: AIC=155.71
## cbind(dataChenilleUseful$pred, dataChenilleUseful$tot_chenilles -
## dataChenilleUseful$pred) ~ F_Elev.mean + F_Elev.kurtosis +
## F_Int.maximum + F_Canopy.relief.ratio + myTableScale$VCI
##
##           Df Deviance   AIC
## <none>           69.035 155.71
## - F_Elev.kurtosis     1   76.818 161.49
## - F_Int.maximum       1   77.645 162.32
## - myTableScale$VCI    1   81.374 166.04
## - F_Canopy.relief.ratio 1   82.026 166.70
## - F_Elev.mean         1  100.165 184.84
##
## Call:  glm(formula = cbind(dataChenilleUseful$pred,
```

```

dataChenilleUseful$tot_chenilles -
##      dataChenilleUseful$pred) ~ F_Elev.mean + F_Elev.kurtosis +
##      F_Int.maximum + F_Canopy.relief.ratio + myTableScale$VCI,
##      family = "binomial", data = dataMetricLidScale)
##
## Coefficients:
##          (Intercept)          F_Elev.mean          F_Elev.kurtosis
##          -2.0359          -1.0314          0.4296
##          F_Int.maximum  F_Canopy.relief.ratio  myTableScale$VCI
##          0.3434          0.5200          -0.6169
##
## Degrees of Freedom: 27 Total (i.e. Null);  22 Residual
## Null Deviance:      110.8
## Residual Deviance: 69.03      AIC: 155.7

# Sélection du meilleur modèle (AIC le plus faible)
mod_select = glm(formula = cbind(dataChenilleUseful$pred,
dataChenilleUseful$tot_chenilles - dataChenilleUseful$pred) ~
                F_Elev.mean + F_Elev.kurtosis +
                F_Int.maximum + F_Canopy.relief.ratio + myTableScale$VCI,
                family = "binomial", data = dataMetricLidScale)

# Nous pouvons maintenant obtenir tous Les paramètres du modèle :
summary(mod_select)

## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -2.0359    0.1069 -19.038 < 2e-16 ***
## F_Elev.mean    -1.0314    0.1855  -5.561 2.68e-08 ***
## F_Elev.kurtosis  0.4296    0.1563   2.748 0.005995 **
## F_Int.maximum   0.3434    0.1153   2.979 0.002897 **
## F_Canopy.relief.ratio 0.5200    0.1462   3.557 0.000375 ***
## VCI            -0.6169    0.1769  -3.487 0.000489 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 110.807  on 27  degrees of freedom
## Residual deviance:  69.035  on 22  degrees of freedom
## AIC: 155.71
##
## Number of Fisher Scoring iterations: 5

# Analyse de La deviance du modèle (pseudo R²)
D2 <- (mod_select>null.deviance - mod_select$deviance) /
mod_select>null.deviance
D2

## [1] 0.3769801

```

Dans ce cas de prédation des défoliateurs, l'application du script montre que la procédure de sélection des variables par AIC ne sélectionne que 5 variables LiDAR (parmi les 10 injectées initialement) : la hauteur moyenne, le kurtosis de la distribution de hauteur, l'intensité maximale, le relief de la canopée et l'indice de complexité verticale. De plus, l'effet de chacune de ces variables est très significatif avec une p-value entre 0 et 0.001 (notre seuil alpha étant fixé à 0.05). Les effets de ces variables n'ont pas tous le même sens. Deux variables ont un effet négatif sur le taux de prédation tandis que les autres ont un effet positif. Cela signifie par exemple que la prédation des chenilles augmente lorsque la hauteur des arbres est moindre et diminue lorsque la structure forestière présente peu de strates verticales (VCI faible en valeur absolue). Enfin, nous obtenons une qualité d'ajustement exprimée par la déviance expliquée (écart entre le modèle calculé et le modèle dit saturé, qui s'ajusterait parfaitement aux données, D2 dans le script). Elle vaut ici environ 0.38 ce qui signifie que 38% du taux de prédation s'explique par ces cinq variables LiDAR.

L'ensemble des résultats obtenus pour chaque modèle (déviance expliquée, variables significatives, sens des effets) est résumé dans le tableau 7 (les numéros de modèles correspondent à ceux donnés dans le tableau 5 section II partie Modélisation statistique par GLM).

Tableau 7 Résultats de chaque modèle GLM obtenu : déviance, variables significatives, taux de significativité, sens des effets.

aOisCan : Abondance en oiseaux s'alimentant dans la canopée
aOisIns : Abondance en oiseaux insectivores
cov2 : Couverture de canopée avec une coupe de hauteur à 2m
cov7 : Couverture de canopée avec une coupe de hauteur à 7m
faibl : Proportion de 1ers retours de faible intensité par rapport au total de 1ers retours
hMax : Hauteur maximale
hMoy : Hauteur moyenne
intMax : Intensité maximale
invD : Inverse de la densité de canopée avec une coupe de hauteur à 2m
Kur : Coefficient d'aplatissement (Kurtosis)
nbR : Nombre de retours supérieurs à 3m
NDVI_moy : Moyenne du NDVI
NDVI_var : Variance du NDVI
notAnc : Note IBP concernant l'ancienneté du bois
notHabaq : Note IBP concernant le nombre de types de milieux aquatiques présents dans le bois
P01 : Percentile 1%
relCa : Relief de la canopée $\left(\frac{H_{moy} - H_{min}}{H_{max} - H_{min}}\right)$
rOisIns : Richesse en oiseaux insectivores
rOisCan : Richesse en oiseaux s'alimentant dans la canopée
Ske : Coefficient de dissymétrie (Skewness)
varl : Variance de l'intensité des 1^{ers} retours
VCI : Indice de complexité verticale
Codes de significativité (p-value) : 0 '****' ; 0.001 '***' ; 0.01 '**' ; 0.05 '.'

Modèle n°	Déviance	Variables significatives	Significativité	Effet
1	0.27	VCI	.	-
2	0.33	VCI	*	-
3	0.58	Kur	*	+
		intMax	.	-
		cov2	.	+
		VCI	**	-

		NDVI_var	*	+
4	0.15	hMoy Kur	* *	+ -
5	0.14	hMax Kur	* .	+ -
6	0.37	hMoy Kur notAnc notHabaq	* ** * **	+ - + +
7	0.38	hMoy Kur intMax relCa VCI	*** ** ** *** ***	- + + + -
8	0.40	hMoy Kur nbR relCa faibl VCI	*** * ** *** ** ***	- + + + - -
9	0.42	hMoy Kur intMax relCa VCI NDVI_moy	*** ** ** *** *** *	- + + + - -
10	0.62	hMoy Kur intMax invD nbR relCa faibl VCI aOisIns	*** * * *** ** *** * *** ***	- + + - + + - - -
11	0.63	hMoy intMax nbR relCa VCI NDVI_moy rOisIns	** *** ** *** *** * ***	- + + + - - -
12	0.55	hMoy Kur invD nbR relCa faibl VCI aOisCan	*** *** ** . *** * *** ***	- + - + + - - -
13	0.56	hMoy Kur nbR relCa VCI NDVI_moy NDVI_var rOisCan	*** ** ** ** *** . . **	- + + + - - + -
14	0.16	hMoy Ske P01 Cov7	*** ** *** ***	- - + -
15	0.43	hMax Kur	* .	- -

		relCa	**	-
16	0.17	varl	.	-
17	0.39	hMax	**	+
		Kur	.	-
		intMax	*	+
		relCa	.	+
18	0.31	hMoy	*	+
		intMax	.	+
19	0.49	hMoy	*	+
		Kur	**	-
		intMax	**	+
		VCI	*	+
20	Non significatif			
21	0.28	hMax	.	+
		hMoy	.	-
		VCI	.	-
22	Non significatif			
23				
24				
25				
26	0.28	hMoy	**	+
		Kur	.	-

Différentes variables LiDAR sont significatives. Nous tenterons d'expliquer le lien entre ces variables exprimant principalement la structure 3D de la végétation forestière et les variables réponses dans la section suivante.

V. Discussion

La variance expliquée par les modèles GLM varie entre 14 et 63%. En matière d'écologie de nombreux paramètres peuvent expliquer la présence ou la diversité des espèces (par exemple les ressources, la compétition entre espèces etc.). Les caractéristiques de l'habitat ne sont donc qu'un facteur parmi d'autres, c'est pourquoi les résultats sont déjà intéressants. De nombreuses variables LiDAR sont sélectionnées dont certaines se retrouvent dans beaucoup de modèles comme par exemple les hauteurs moyenne et maximale, le coefficient d'aplatissement de la distribution des hauteurs (Kurtosis) et le VCI.

Concernant l'indice de complexité verticale des interrogations demeures. Comme nous l'avons expliqué dans la partie cas d'étude nous avons réalisé une sortie terrain afin de mieux se rendre compte de la réalité physique des variables LiDAR extraites. La figure 23 (autres exemples en annexe 3) réalisée à partir de captures d'écran de FugroViewer, nous permet de visualiser la donnée et d'ainsi mieux comprendre les variables LiDAR à notre disposition (ici le VCI). Nous pouvons en effet visualiser les différences dans les nuages de points pour des valeurs extrêmes d'une métrique particulière, et la sortie terrain permet de vérifier la cohérence de ces résultats avec les disparités visibles sur le terrain (Figure 24, autres exemples en annexe 3).

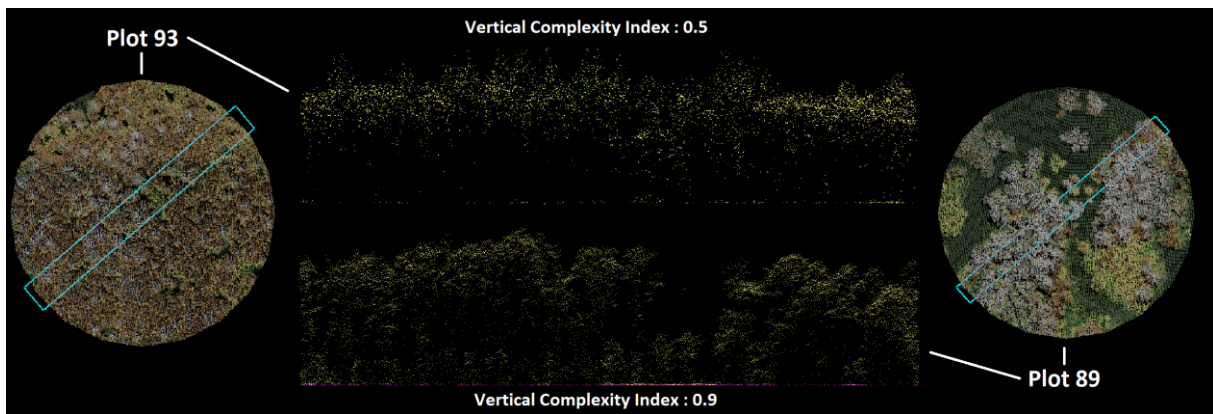


Figure 23 Visualisation des données LiDAR pour deux plots présentant des valeurs extrêmes de VCI.

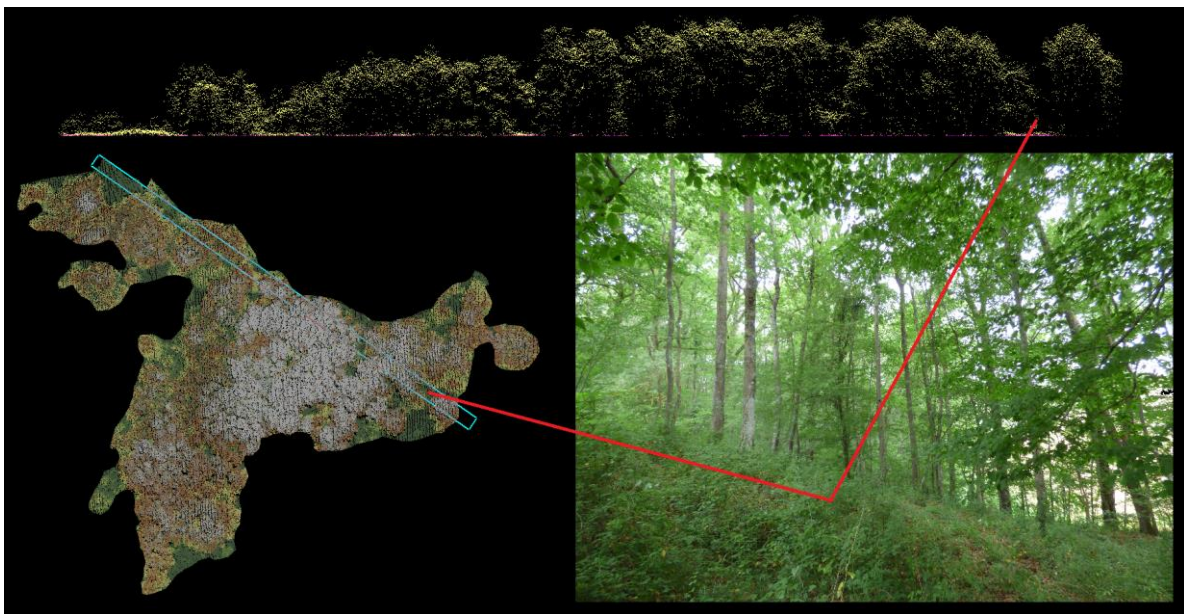


Figure 24 Photographie terrain et coupe verticale du nuage de points.

Il se trouve que les différences terrain entre des bois et plots présentant des valeurs extrêmes de VCI n'étaient pas exactement celles à quoi on s'attendait. En effet, nous pensions décrire avec cet indice le degré d'encombrement d'une zone mais il semble davantage caractériser la présence de plusieurs strates dans des milieux plutôt ouverts. Le faisceau LiDAR ayant du mal à pénétrer la canopée dense dans le cas de nos levés, la densité des strates inférieures de la canopée se distingue difficilement (peu de retours au-delà du premier écho).

Il est à présent intéressant de chercher à comprendre pourquoi les variables LiDAR significatives expliquent les variables réponses. Nous avons jusqu'ici adopté une approche très exploratoire en limitant les hypothèses sur les relations possibles ou pas entre variables. La compréhension des effets est maintenant nécessaire. Pour cela, nous regarderons les modèles présentant une

déviante expliquée supérieure à 49% (modèles n°3, 10, 11, 12, 13, 19). Ce choix est réalisé en tenant compte du nombre limite de pages.

Le modèle 3 cherche à expliquer l'IBP des relevés MUESLI (échelle des bois) à partir des variables LiDAR, de la richesse en oiseaux qui nichent dans les cavités et de la surface des bois. Les variables LiDAR sont introduites dans le modèle car elles sont susceptibles d'expliquer les facteurs, permettant de calculer l'IBP, liés au peuplement et à la gestion forestière. La structure verticale de la végétation qui est notée en fonction du nombre de strates présentes (parmi les strates herbacée, ligneuse basse, intermédiaire et haute) pourrait être expliquée par le VCI. La note caractérisant la surface de milieux ouverts au sein de la forêt pourrait, elle, se rapprocher de la couverture ou de la transparence de canopée. Cependant, comme nous avons pu le dire, dans une approche exploratoire nous ne restreignons pas la modélisation à ces seules hypothèses de liens et utilisons l'ensemble des variables LiDAR à disposition. La richesse en oiseaux qui nichent dans les cavités est utilisée en raison du facteur reflétant la présence d'arbres porteurs de microhabitats dans le calcul de l'IBP. En effet, la présence d'oiseaux nichant dans les cavités est susceptible de révéler la présence d'un certain nombre de cavités et donc de microhabitats. La surface des bois est, elle, utilisée car nous savons que l'abondance et la richesse en oiseaux augmente avec la taille de la forêt. Il est donc probable que cette variable puisse expliquer l'IBP.

Cinq variables issues de la télédétection sont significatives dans ce modèle : le kurtosis (aplatissement de la distribution), l'intensité maximale, la couverture de canopée avec une coupe de hauteur à 2m, le VCI et la variance du NDVI. A noter que la richesse en oiseaux nichant dans les cavités est aussi sélectionnée dans le modèle mais n'est pas significative. La note d'IBP semble donc augmenter dans des forêts encombrées et présentant du bois mort. En effet, elle augmente avec l'augmentation de la couverture de canopée, l'augmentation de la complexité verticale c'est-à-dire la présence de davantage de strates (effet négatif mais valeur de VCI négatives) et l'augmentation du kurtosis donc présentant une distribution peu étirée de hauteurs. La note IBP augmente aussi avec la diminution de l'intensité maximale des échos LiDAR. Or, Wing et al. (2015) ont montré que l'intensité LiDAR pouvait différer entre les arbres sains et les arbres morts, ces derniers renvoyant une intensité LiDAR plus faible. L'un des facteurs de la note IBP étant la présence de bois mort sur pied (et au sol), l'effet obtenu avec la variable d'intensité pourrait s'expliquer de cette façon, en accord avec l'observation de Wing et al. (2015).

Les variables extraites expliquent 58% de l'IBP dans ce cas. Cependant, les résultats obtenus sont plus faibles en utilisant seulement les variables LiDAR (27% à l'échelle du bois, 14% pour les plots). La faible déviante expliquée peut être due à la faible prise en compte directe de la

structure, notamment verticale, des forêts dans le calcul de l'IBP (structure verticale de la végétation et milieux ouverts seulement). De plus, de nombreux facteurs de l'IBP se basent sur un nombre d'arbres (gros bois mort sur pied, au sol, très gros bois vivant, arbres vivants porteurs de microhabitats). Une segmentation permettant d'obtenir des données LiDAR à l'échelle des arbres est donc nécessaire pour se rapprocher de la méthode de calcul de l'IBP.

Le modèle 19 indique que la richesse en oiseaux nichant dans les cavités augmente dans les forêts présentant des arbres hauts (hauteur moyenne) et vivants (intensité maximale) tout en présentant une homogénéité verticale c'est-à-dire peu de strates végétales (VCI). D'un point de vue écologique, la présence de cavités dans une forêt est notamment liée à : sa maturité, sa gestion (cavités dues aux coupes ou a contrario bois mort accumulé par le laisser-faire) et des événements extérieurs (par exemple des cassures de branches dues à une tempête). La présence d'arbres hauts pourrait être révélatrice d'arbres adultes (matures) bien développés et abritant donc davantage de cavités. La gestion forestière semble aussi avoir une influence non négligeable sur la présence de cavité comme en atteste l'augmentation due au peu d'encombrement des forêts et à la présence d'arbres vivants. Les résultats ne paraissent donc pas aberrants mais restent à confirmer.

Les autres modèles (n°10 à 13) cherchent à expliquer le taux de prédation des défoliateurs à partir des variables LiDAR et de la richesse et l'abondance en oiseaux insectivores ou s'alimentant dans la canopée. La relation obtenue entre le taux de prédation et la richesse ou l'abondance en oiseaux est négative ce qui signifie que le taux de prédation diminue quand le nombre et la diversité en oiseaux insectivores augmente. Ce résultat est surprenant et inverse à l'attendu. On peut effectivement penser que lorsque qu'il y a plus d'oiseaux insectivores, il y a plus de prédation. Il se trouve que le groupe des oiseaux insectivores regroupe des espèces assez différentes qui ne s'alimentent pas toutes dans la canopée. Certains se nourrissent au sol (strate herbacée et buissonnante). Ces derniers n'auront pas d'effet sur la prédation. Par conséquent, s'ils sont plus nombreux, cette relation inverse pourrait apparaître. Il serait donc intéressant d'affiner le regroupement des oiseaux insectivores en ne sélectionnant que ceux qui s'alimentent dans la canopée, là où les chenilles sont présentes. Cette spécialisation du groupe d'insectivores sera à prévoir pour la suite. Une autre explication de ce résultat est la potentielle interaction entre la composition et/ou la diversité des communautés d'oiseaux et l'environnement forestier. La prédation devrait diminuer avec la taille de la forêt (les chenilles étant plus difficile à trouver) alors que la diversité devrait, elle, augmenter avec cette taille (relation aire-espèce en écologie). La surface seule ne semble pas avoir d'effet sur la prédation car elle n'est pas sélectionnée mais

peut-être que l'interaction entre la surface et la richesse (ou abondance) a, elle, un effet ; ce qu'il serait intéressant de vérifier pour la suite.

En outre, nous avons différencié la richesse en oiseaux en construisant différentes catégories selon leurs habitudes d'alimentation et leur lieu de nidification. Cette différenciation en groupes fonctionnels a pu être réalisée à partir d'une base de traits des espèces fournie par les écologues du laboratoire. Ces modèles (n°19 à 25) ont été créés pour se rapprocher des travaux de Lesak et al. (2011), qui ont aussi cherché à modéliser la richesse en oiseaux forestiers à partir de mesures LiDAR. Ils réussissent à obtenir des modèles significatifs à partir de quelques variables LiDAR (percentiles de hauteur, densités de points par tranches de hauteur en particulier) alors que les nôtres ne le sont pas. Ces résultats sont cependant à nuancer par la faible valeur de R^2 de leurs modèles (aux alentours de 10% de déviance expliquée). Ces différences peuvent s'expliquer par la densité de points des levés LiDAR, les méthodes statistiques utilisées, le très faible nombre d'observations à notre disposition ou bien encore, l'échelle d'analyse et la nature des forêts étudiées.

De façon générale, ces premiers résultats obtenus doivent aussi être traités avec recul pour plusieurs raisons. Tout d'abord en ce qui concerne les données IBP correspondant aux relevés hors MUESLI, ils ont été réalisés en 2011. Un décalage entre les observations terrain et les mesures LiDAR peut donc influencer les résultats, les forêts ayant évolué durant cette période et leur gestion pouvant avoir été modifiée. Il serait intéressant d'étudier si ce décalage explique les différences notables dans les résultats obtenus à l'échelle des bois et celle des plots (27% de déviance expliquée pour la première contre 14% pour la seconde en utilisant seulement les variables LiDAR). Ensuite, nous disposons de peu d'observations pour construire nos modèles : 28 bois et 104 plots pour de nombreuses variables LiDAR. Le nombre de plots peut sembler plus important mais il est à relativiser car chaque bois contient 3 à 4 plots ce qui entraîne une autocorrélation spatiale comme évoquée en amont. De plus, les bois de la zone ne présentent pas assez de disparités tout comme les variables LiDAR qui en sont extraites. Par conséquent, il convient à ce stade d'être prudent avant de généraliser et appliquer les modèles sur de larges étendues. D'un point de vue statistique, le faible nombre d'observations engendre également des difficultés pour choisir la distribution de certaines variables réponses. Pour certains groupes d'oiseaux par exemple, la richesse (correspondant à un comptage) aurait dû faire appel, en théorie, à une famille poissonnienne mais le test de Shapiro nous a conduits à conserver un modèle Gaussien.

Pour ces raisons, les résultats de la spatialisation doivent être validés. En effet, certains modèles permettent d'expliquer une grande partie de la variable réponse mais les remarques faites précédemment montrent qu'au mieux seule une valeur relative peut être obtenue. Nous avons testé cette spatialisation à partir du modèle construit n°2 permettant d'expliquer l'IBP à partir des variables LiDAR et du NDVI. Dans ce modèle seul le VCI est significatif mais le coefficient d'aplatissement de la distribution de hauteur, l'intensité maximale, la couverture de canopée avec un seuil de hauteur supérieur à 2m et la variance du NDVI sont aussi utilisés. Cette prédiction a été réalisée sur 11 bois de la zone (Figure 25). Nous pouvons noter de fortes disparités dans ces bois mais une validation des résultats est donc nécessaire. Comme nous le verrons en conclusion et dans les perspectives, il serait intéressant d'obtenir d'autres mesures terrain afin d'évaluer le pouvoir prédictif sur de larges zones des modèles créés.

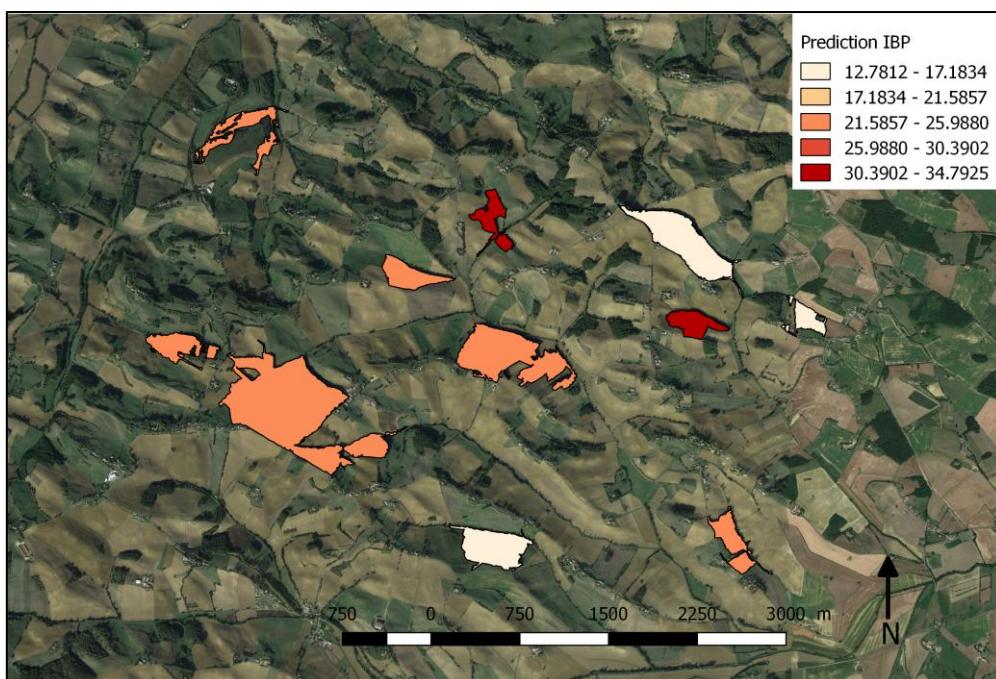


Figure 25 Prédiction de l'IBP à partir de variables LiDAR et de la variance du NDVI (correspondant au modèle créé n°2) sur 11 bois de la zone des levés.

5. EXTRACTION AUTOMATIQUE DES CIMES ET COURONNES DES ARBRES

I. Intérêts et méthodes existantes

Comme nous l'avons présenté dans la partie sur l'état de l'art, les variables LiDAR peuvent être extraites à différentes échelles. Une de ces échelles est celle de l'arbre d'où la nécessité d'extraire les cimes et les couronnes des arbres. Cette action étant fastidieuse à la main sur de grandes zones, il apparaît indispensable d'automatiser le processus.

Les recherches bibliographiques menées sur ce sujet font état de trois démarches principales pour effectuer cette extraction. Elle peut être réalisée soit directement à partir du nuage de points, à l'aide du MNH ou encore grâce à une méthode mixant les deux produits précédents. Zhen et al. (2016) dressent un état des différentes méthodes utilisées dans les publications. Le sujet de l'utilisation des données LiDAR pour l'extraction automatique des cimes et couronnes d'arbres revient dans différentes communautés : télédétection, SIG, foresterie et informatique. Des technologies passives comme l'imagerie hyperspectrale sont aussi utilisées pour effectuer l'extraction des arbres individuels mais nous n'en parlerons pas ici. A noter qu'au cours des 25 dernières années les articles traitant des données actives pour traiter ces questions ont nettement augmenté alors que ceux sur les technologies passives ont tendance à stagner.

Les méthodes basées sur l'analyse d'un raster (MNH) sont les plus nombreuses suivies par celles concentrées sur le nuage de points alors que celles combinant les deux sont les moins nombreuses. Concernant la première méthode elle possède l'avantage d'être bien développée, facile d'utilisation et d'amélioration mais en revanche manque de précision et peut causer des erreurs. Il s'agit par exemple d'une détection des cimes par détection des maxima locaux (e.g. Hyypä et al. 2001) ou une binarisation de l'image suivie d'une détection des couronnes par la méthode des bassins-versants (aussi appelée ligne de partage des eaux).

Les méthodes construites sur l'analyse du nuage de points permettant quant à elle de mieux utiliser l'information 3D et donc de mieux refléter la structure de la canopée en détectant non seulement le haut de la canopée mais aussi les sous-strates ; ce que ne permet pas la méthode précédente se basant sur le MNH. Elles sont en revanche plus compliquées à implémenter et très dépendantes de la qualité des données LiDAR et plus particulièrement de la densité du nuage de points. Nous retrouvons ainsi le partitionnement par k-moyennes (*k-means clustering*) et la

transformation du nuage de points en voxels¹² (e.g. Wang et al. 2008) qui permettent de regrouper les points issus d'un même arbre par analyse de leur voisinage. Nous pouvons alors obtenir une segmentation par arbres individuels reflétant différentes strates dans la forêt.

Enfin, il existe des méthodes combinant à la fois le MNH (raster) et le nuage de points (en plus de connaissances a priori). Elles permettent d'accumuler les données par la prise en compte de différentes sources (télédétection et analyse spatiale). Nous pouvons ainsi prendre en compte à la fois la hauteur des objets mais aussi leur composition spectrale ou bien des informations extérieures comme le contexte historique. Ce dernier point représente aussi un désavantage à la méthode en raison de la possible difficulté à agglomérer l'ensemble de ces informations, issues de sources hétérogènes. Cette méthode peut être un des exemples cité précédemment comme la détection des maximas locaux en ajoutant une information a priori sur la taille moyenne des couronnes (e.g. Zhen et al. 2015).

Pour aller plus loin dans la détection individuelle des arbres, des méthodes existent pour non pas seulement détecter la cime et la couronne mais pour reconstruire la forme de l'arbre. Il s'agit par exemple des méthodes se basant sur les notions mathématiques de la transformée de Hough (e.g. Van Leeuwen et al. 2004), des superquadratiques (e.g. Weinacker et al. 2010) ou bien encore de l'enveloppe convexe. Nous pouvons ainsi récupérer le profil vertical de l'arbre ou mieux estimer leur feuillage mais les algorithmes à implémenter sont lourds à mettre en œuvre, nécessitent des relevés terrains précis pour valider les calculs et peuvent dépendre de la qualité de la segmentation initiale.

II. Démarche adoptée et mise en œuvre

Nous avons voulu détecter les cimes des arbres car cela pourrait servir, par la suite, à ajouter de nouvelles variables LiDAR dans les modèles créés. Ces métriques pourraient alors être pertinentes pour notre étude étant donné que certains indicateurs de biodiversité se basent sur des comptages d'arbres (l'IBP tient notamment compte du nombre de très gros bois au sein d'une forêt) d'où la nécessité de les individualiser.

Comme nous l'avons vu dans l'organisation du travail, le temps consacré à la segmentation a été nettement moins important que celui pour la modélisation statistique. C'est pourquoi nous avons choisi de tester une approche simple, déjà opérationnelle, et d'en mesurer sa performance sur un peuplement régulier (plantation de peupliers alignés, Figure 26). La méthode choisie est celle

¹² Pixels 3D

basée sur l'analyse du MNH : détection des maxima locaux avec information sur le diamètre des arbres et méthode de ligne de partage des eaux.

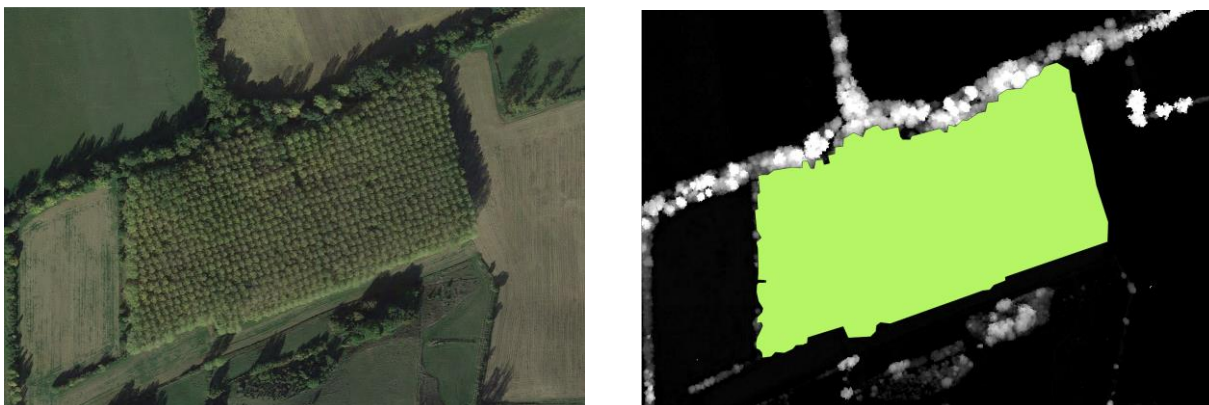


Figure 26 Photographie aérienne de la zone d'étude, issue de Google maps (à gauche) et entité vectorielle dessinée à partir du MNH (à droite).

Ce choix présente l'avantage de pouvoir évaluer correctement le résultat obtenu. En effet, les cimes des arbres sont bien discernables dans ce peuplement et elles ont été identifiées manuellement sous SIG (QGIS) à partir du MNH ; de façon à créer un jeu de validation complet (couche de points). Un total de 1040 cimes d'arbres a été identifié dans cette plantation (Figure 27).

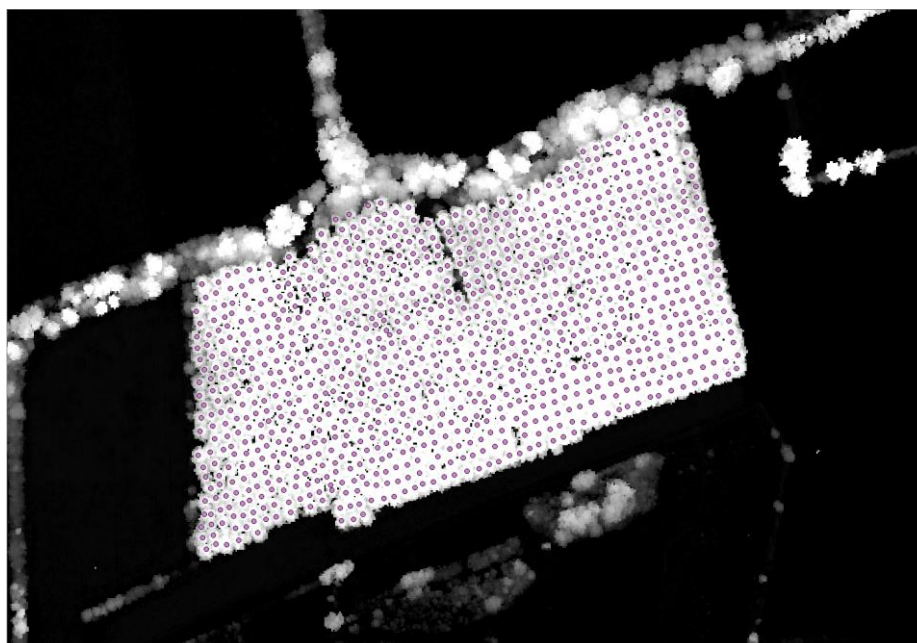


Figure 27 Couche vectorielle créée avec l'emplacement de chaque cime.

Une fois cet échantillon de validation créé, le MNH issu des données LiDAR (résolution d'1m) a été lissé à l'aide d'un filtre gaussien de taille de voisinage variable avant de détecter les cimes. Des

filtres gaussien 3*3 (rayon de recherche = 1) et 5*5 (rayon de recherche = 2) ont été utilisés à l'aide de la fonction de filtrage gaussien d'OTB (Orfeo ToolBox). La détection des cimes a aussi été réalisée sans filtrage du raster afin de comparer les différents résultats.

Le script Python¹³ développé par Matthew Parkan de l'Ecole Polytechnique Fédérale de Lausanne (EPFL) basé sur la méthode des maxima locaux pour détecter le sommet des arbres à partir d'un MNH a d'abord été utilisé. Deux paramètres doivent être fixés : la hauteur minimale des arbres à détecter ainsi que la fenêtre de voisinage c'est-à-dire le rayon dans lequel le maximum local est recherché. Le premier paramètre a été maintenu à 3m correspondant au seuil utilisé pour le calcul des variables LiDAR avec les logiciels LAsTools et FUSION/LDV. Le second a en revanche fait l'objet de différents tests, variant de 2.5m à 5m avec un pas de 0.5m. Ce choix a été fait en raison du rayon moyen des arbres de la zone d'intérêt qui se situe dans cette fourchette de valeurs. Cette méthode permet donc d'obtenir une couche de points correspondant aux cimes des arbres.

L'algorithme « TreeSeg » de FUSION/LDV a été utilisé en parallèle pour comparaison et extraction des couronnes. Il applique une segmentation par ligne de partage des eaux sur un MNH pour construire des bassins correspondant aux couronnes des arbres. Le résultat inclut, en plus des couronnes, les cimes des arbres correspondant aux points les plus hauts des bassins calculés. A noter que la hauteur minimale des objets est là aussi fixée à 3m.

L'ensemble des résultats obtenus par les deux algorithmes et avec les différents paramètres sont alors évalués à l'aide de l'échantillon de validation. Cela permet d'estimer la précision des algorithmes dans ce cas particulier de peupliers alignés.

III. Résultats

Les algorithmes lancés permettent d'obtenir des couches de points visualisables sur QGIS par exemple. Les figures suivantes présentent les résultats obtenus d'une part avec le script Python (cimes par maxima locaux, figures 28 à 30) et d'autre part avec FUSION/LDV (couronnes et cimes, figures 31 à 33). La zone d'intérêt est représentée en orange pâle et les sommets des arbres placés à la main sont en noir tandis que les cimes calculées sont en rouge. Sur les résultats de la fonction « TreeSeg » les couronnes sont aussi représentées, en blanc et détourées en noir. A noter que concernant le script Python sont reportés dans cette partie seulement les résultats obtenus avec un filtre gaussien 3*3 et un rayon de recherche compris entre 3.5m et 4.5m. L'ensemble des résultats est cependant fourni en annexe 6.

¹³ <https://github.com/mparkan/sylva>

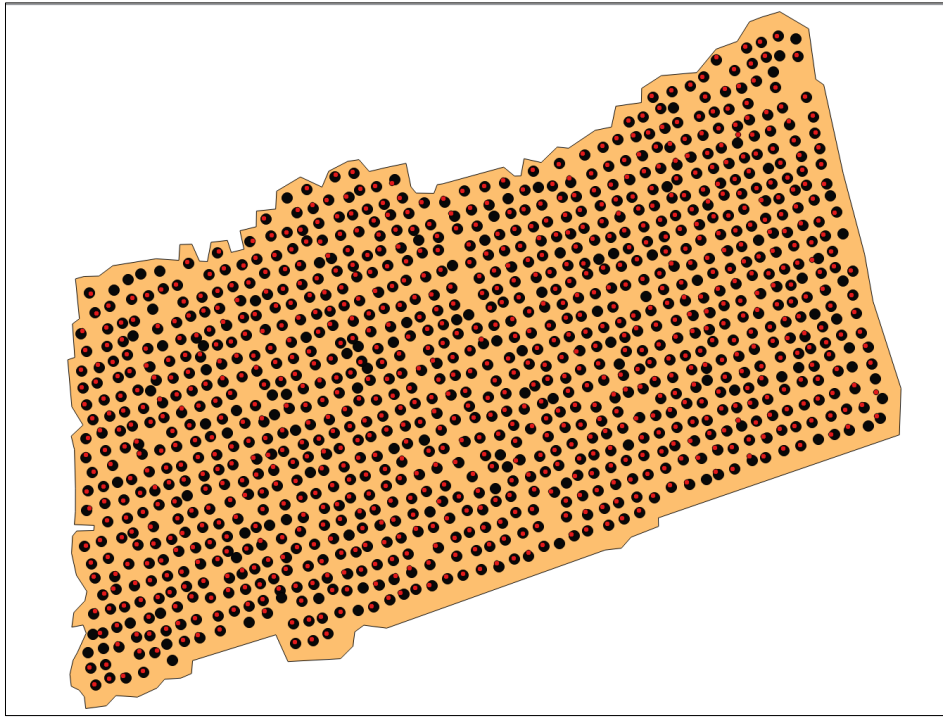


Figure 28 Résultat avec la méthode des maxima locaux (script Python) et un filtrage gaussien 3*3 rayon de recherche de 3.5m.

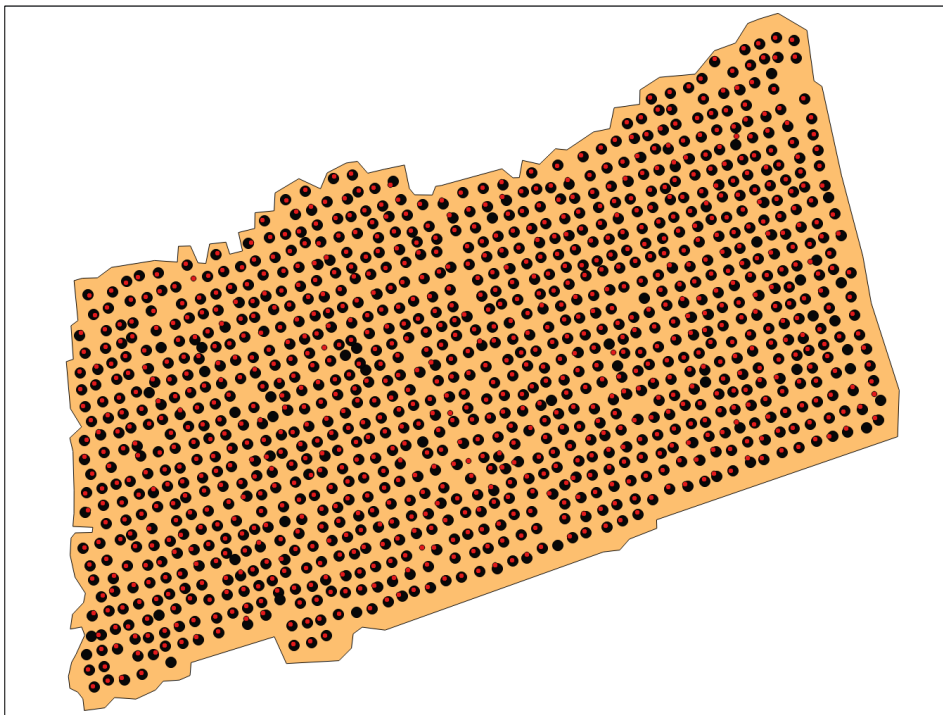


Figure 29 Résultat avec la méthode des maxima locaux (script Python) et un filtrage gaussien 3*3 rayon de recherche de 4m.

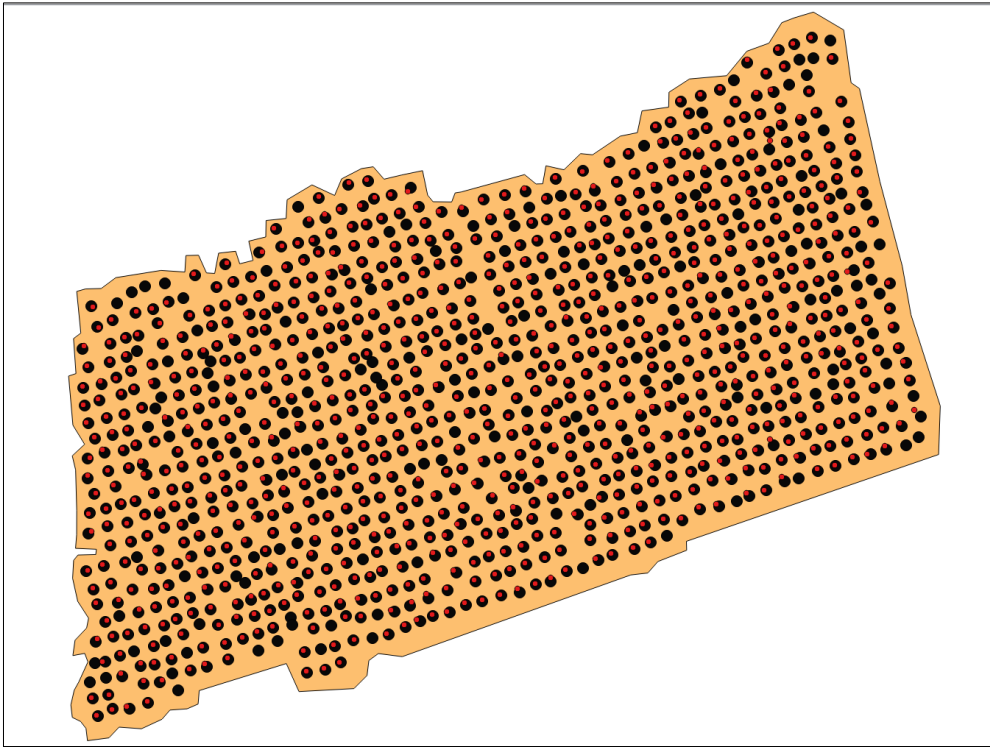


Figure 30 Résultat avec la méthode des maxima locaux (script Python) et un filtrage gaussien 3*3 rayon de recherche de 4.5m.



Figure 31 Résultat avec FUSION et un filtrage gaussien 3*3.

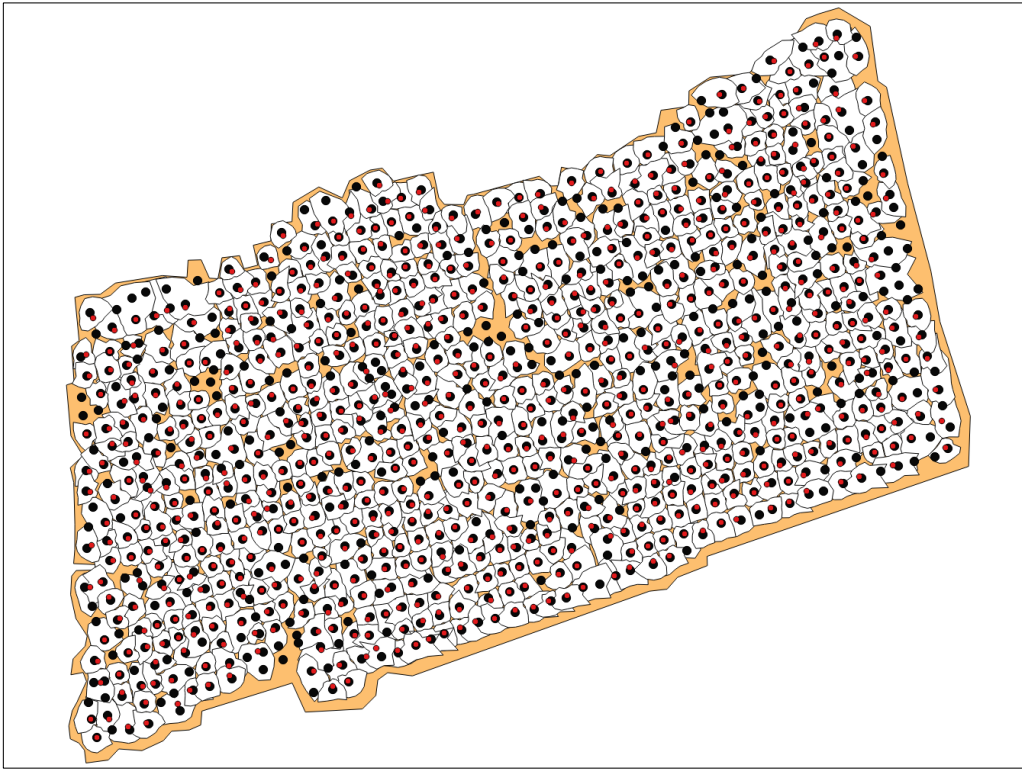


Figure 32 Résultat avec FUSION et un filtrage gaussien 5*5.

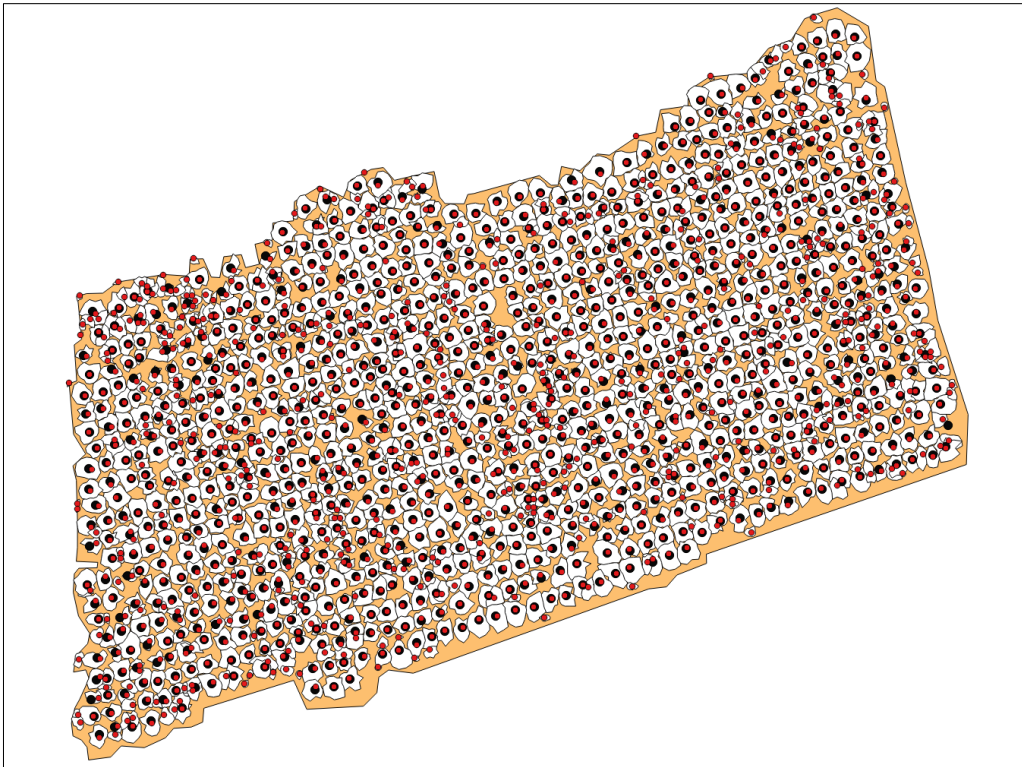


Figure 33 Résultat avec FUSION sans filtrage.

Sur les figures précédentes nous pouvons déjà remarquer l'influence des différents paramètres ainsi que des méthodes utilisées. Afin de mieux comparer ces résultats de segmentation, nous avons décidé de calculer la distance entre les cimes détectées par les algorithmes et celles permettant la validation. Une fois cette distance calculée nous supprimons les distances supérieures au rayon de recherche divisé par 2 ou bien supérieures à 2m. Autrement dit, si la distance entre les points issus de chaque couche est trop importante nous considérons que la cime n'a pas été détectée ou est fautive. Ce choix se justifie par la résolution du MNH. En effet, celle-ci étant d'1m le seuil fixé représente environ 2 pixels et reste inférieur au rayon des arbres. Nous pouvons donc considérer que l'arbre détecté est le bon. Nous calculons alors la distance euclidienne moyenne entre les cimes détectées par les algorithmes et celles de l'échantillon de validation, ainsi que les taux d'accord, les erreurs d'omission (déficits) et de commission (excédents). Cela est fait pour chaque paramétrage (voir tableaux 8 à 11).

Tableau 8 Résultats de l'extraction par maxima locaux pour un filtrage gaussien 5*5 : distance euclidienne moyenne par rapport à la couche de validation, nombre de cimes valides parmi celles obtenues, erreur d'omission et de commission pour chaque rayon de recherche et par rapport à la distance maximale de validation d'une cime.

	Filtrage gaussien R = 2					
Rayon de recherche (m)	2,5	3	3,5	4	4,5	5
	Suppression des distances > (rayon de recherche/2) m					
Distance moyenne (m)	0,75	0,87	0,99	1,06	1,07	1,08
Taux d'accord (%)	58	72	88	94	98	99
Omission (%)	55	51	38	45	45	56
Commission (%)	33	20	8	4	1	1
	Suppression des distances > 2m					
Distance moyenne (m)	1,06	1,06	1,04	1,06	1,04	1,05
Taux d'accord (%)	93	94	94	94	95	96
Omission (%)	28	35	35	45	47	58
Commission (%)	5	4	4	4	3	2

Tableau 9 Résultats de l'extraction par maxima locaux pour un filtrage gaussien 3*3 : distance euclidienne moyenne par rapport à la couche de validation, nombre de cimes valides parmi celles obtenues, erreur d'omission et de commission pour chaque rayon de recherche et par rapport à la distance maximale de validation d'une cime.

	Filtrage gaussien R = 1					
Rayon de recherche (m)	2,5	3	3,5	4	4,5	5
	Suppression des distances > (rayon de recherche/2) m					
Distance moyenne (m)	0,75	0,82	0,85	0,86	0,85	0,86
Taux d'accord (%)	82	88	98	99	100	100
Omission (%)	17	10	5	10	15	27
Commission (%)	18	7	2	1	0,5	0,5
	Suppression des distances > 2m					

Distance moyenne (m)	0,86	0,86	0,86	0,86	0,85	0,85
Taux d'accord (%)	97	99	99	100	100	100
Omission (%)	2	4	4	10	15	27
Commission (%)	3	1	1	0,5	0,5	0,5

Tableau 10 Résultats de l'extraction par maxima locaux sans filtrage : distance euclidienne moyenne par rapport à la couche de validation, nombre de cimes valides parmi celles obtenues, erreur d'omission et de commission pour chaque rayon de recherche et par rapport à la distance maximale de validation d'une cime.

	Sans filtrage					
Rayon de recherche (m)	2,5	3	3,5	4	4,5	5
	Suppression des distances > (rayon de recherche/2) m					
Distance moyenne (m)	0,72	0,76	0,80	0,81	0,81	0,82
Taux d'accord (%)	71	90	94	98	99	100
Omission (%)	14	9	5	9	13	26
Commission (%)	36	10	7	2	1	0,5
	Suppression des distances > 2m					
Distance moyenne (m)	0,84	0,82	0,82	0,82	0,81	0,81
Taux d'accord (%)	83	95	95	98	98	99
Omission (%)	0	3	3	9	13	27
Commission (%)	21	5	5	2	1	1

Tableau 11 Résultats de l'extraction par ligne de partage des eaux : distance euclidienne moyenne par rapport à la couche de validation, nombre de cimes valides parmi celles obtenues, erreur d'omission et de commission pour chaque filtrage.

	Algorithme « TreeSeg » – FUSION/LDV		
Filtrage	R = 2	R = 1	Sans
	Suppression des distances > 2m		
Distance moyenne (m)	0,88	0,59	1,36
Taux d'accord (%)	96	97	74
Omission (%)	31	1	0
Commission (%)	3	3	36

Nous remarquons tout d'abord que la distance moyenne entre les points est souvent inférieure au mètre. Nous pouvons donc considérer que les cimes détectées sont situées dans le voisinage du pixel défini manuellement comme étant la cime. Les méthodes utilisées permettent donc la localisation précise des cimes, confirmé par le taux d'accord de cimes détectées (nombre de cimes réelles par rapport au nombre de cimes détectées) qui est souvent supérieur à 90%. Le filtrage 5*5 pour la méthode des maxima locaux et le non filtrage pour l'algorithme « TreeSeg » présentent toutefois des valeurs de distances moyennes supérieures au mètre. Ces deux cas de segmentation semblent donc à écarter dans notre cas d'étude. Il reste à présent à évaluer si le

résultat présente des déficits et/ou excédents. En regardant les différentes erreurs d'omission, à rayon de recherche égal pour chaque filtrage, le filtre 5*5 semble une nouvelle fois à écarter. En effet, des centaines d'arbres ne sont pas détectés (entre 28% et 58%). Concernant l'autre filtrage et les résultats sans filtre, l'erreur d'omission a tendance à augmenter avec le rayon de recherche tandis que l'erreur de commission diminue. Cela semble tout à fait logique et il est donc intéressant de trouver le rayon présentant le meilleur compromis afin d'avoir deux erreurs les plus faibles possibles. Cela revient dans notre étude à utiliser un rayon de recherche compris entre 3 et 3,5m. De plus, un filtrage de rayon 1 semble nécessaire pour ne pas surdétecter de sommets sans toutefois en omettre trop. En outre, si nous comparons les deux algorithmes, dans le cas donc d'un filtrage de rayon 1, les résultats sont semblables pour un rayon de recherche bien choisi dans le cas de la méthode des maxima locaux. L'algorithme « TreeSeg » basé sur une segmentation par bassins versants est donc plus efficace dans le cadre de notre zone d'intérêt. Effectivement, pour des erreurs d'omission et de commission identiques il permet de s'affranchir de la recherche préalable du diamètre moyen des arbres pour paramétrer la segmentation. De plus, cette méthode permet d'obtenir les couronnes des arbres en complément des cimes contrairement à la méthode par calcul des maxima locaux.

IV. Discussion

Quelques conclusions ont déjà été mises en avant dans les résultats et il est intéressant ici de réfléchir aux améliorations futures. Tout d'abord, il semble utile de filtrer le MNH avant d'effectuer la segmentation comme c'est préconisé dans divers articles traitant ce sujet (Dufort 2000). Attention toutefois à ne pas le filtrer avec une taille de voisinage trop élevée afin de ne pas accroître considérablement le taux d'omissions suite au lissage.

Ensuite, comme nous l'avons vu la méthode par ligne de partage des eaux est ici la meilleure. Elle permet d'obtenir de bonnes précisions tout en obtenant à la fois les cimes et les couronnes. Cependant, il serait intéressant de tester par la suite une autre méthode. Elle consisterait à utiliser celle de ligne de partage des eaux mais cette fois-ci contrôlée par des marqueurs (i.e. une approche semi-supervisée). Ces marqueurs seraient les maxima locaux et permettraient ainsi de contraindre la recherche des bassins versants afin d'éviter une sur-segmentation (Amiri 2014).

De plus, d'autres méthodes de segmentation sont à tester. Effectivement, nous avons traité ici les méthodes basées sur l'analyse du raster extrait des données LiDAR qui permet seulement de segmenter le niveau supérieur de la canopée. Or tout l'intérêt des données LiDAR réside dans la perception de la structure verticale de la végétation. Il serait donc intéressant d'essayer les

méthodes de segmentation vues dans la section I de cette partie, basées sur l'analyse du nuage de points ou bien encore des informations a priori. Ainsi, les sous-strates pourraient être mises en avant et nous pourrions obtenir, pour des données LiDAR présentant une forte densité, la structure complète de chaque arbre, individuellement.

En outre, l'étude menée ici a été réalisée seulement sur une zone test considérée comme facile à segmenter. En effet, elle présente des arbres alignés d'une même espèce. Ce choix a été fait en raison du faible temps dédié dans ce stage à cette partie mais les méthodes devront par la suite être testées sur la zone complète de la couverture LiDAR. Cela afin d'évaluer les résultats d'une segmentation sur des zones plus complexes, correspondant aux forêts typiques des coteaux de Gascogne, composées principalement de chênes mais mêlant aussi d'autres espèces telles le charme et le merisier.

Une fois cette segmentation réalisée sur l'ensemble de la zone, les résultats obtenus, s'ils sont satisfaisants, pourraient servir à extraire de nouvelles variables LiDAR. En effet, nous obtiendrons ainsi des entités à une nouvelle échelle (celle de l'arbre) où des métriques LiDAR seraient calculées. Comme nous l'avons vu dans la partie état de l'art sur les variables LiDAR usuelles à l'échelle de l'arbre, des modèles allométriques pourraient être mis en place. Cela afin de mieux caractériser les habitats en connaissant par exemple le diamètre des arbres, leur volume ou bien encore la surface terrière. Comme nous l'avons évoqué, l'IBP prend par exemple en compte le nombre de très gros bois d'où la nécessité d'individualiser les arbres.

CONCLUSION

Les résultats obtenus sont encourageants : certaines variables LiDAR permettent d'expliquer significativement entre 14 et 63% des variables de biodiversité étudiées. Au-delà de 40% le pouvoir explicatif du modèle est considéré comme élevé en écologie. Les résultats sont d'autant plus intéressants que les caractéristiques de l'habitat, seul paramètre explicatif ici, ne sont qu'un facteur parmi d'autres permettant d'expliquer la présence ou la diversité des espèces.

Nous pouvons ainsi spatialiser, au moins en valeur relative, certains services écosystémiques sur de larges zones. Cela permet d'obtenir une première idée sur la nature des bois de la zone (connaissance d'entités aux valeurs extrêmes par exemple) préalablement aux relevés terrains. Les résultats de la segmentation sont aussi intéressants sur la zone d'étude simplifiée.

Les travaux effectués lors de ce stage laissent entrevoir des perspectives de recherche pour le futur. Premièrement, il est à présent intéressant d'étendre l'application des algorithmes de détection des cimes et couronnes à l'ensemble des forêts couvertes par les levés LiDAR. De plus, d'autres algorithmes de segmentation pourraient être testés notamment afin d'extraire des informations directement du nuage de points et ainsi dégager de nouvelles variables LiDAR à injecter dans les modélisations statistiques. Cela permettrait de ne pas perdre l'information correspondant aux sous-strates de la végétation.

Deuxièmement, concernant certains indicateurs de biodiversité, ils pourraient être divisés. Nous avons ici différencié la richesse totale en oiseaux de celle propre à différentes catégories ce qui pourrait se faire notamment pour la richesse de la strate herbacée. Nous pourrions différencier les espèces sciaphiles des espèces héliophiles qui, elles, recherchent la lumière et sont donc susceptibles d'être liées, par exemple, à la transparence de la canopée.

Enfin, davantage d'indicateurs pourraient être prédits tout en vérifiant leur validité. Des échanges ont eu lieu en ce sens avec des chercheurs du CEFS (Comportement et Ecologie de la Faune Sauvage) afin par exemple de lier les cartes de prédiction de l'abrutissement par le chevreuil obtenues à partir de la modélisation statistique aux déplacements des chevreuils mesurés par relevés GPS.

D'un point de vue personnel, ce stage de fin d'études effectué au sein du laboratoire Dynafor m'a permis de mieux connaître le fonctionnement d'un laboratoire de recherche. Durant ces mois de stage, j'ai eu des échanges très constructifs et obtenu de nombreux conseils auprès des chercheurs que j'ai côtoyés. Cela m'a permis de préciser mon projet professionnel. J'ai ainsi écarté

l'idée de faire une thèse de doctorat car je me suis rendu compte que je préfère, à partir d'un problème précis et concret, essayer d'apporter la solution la plus optimale possible en fonction des contraintes du projet (temps imparti, ressources à disposition, technologie à utiliser...). Il me semble qu'un travail dans un laboratoire de recherche consiste à creuser beaucoup plus en profondeur une problématique afin d'acquérir une certaine expertise dans le domaine et proposer de nouvelles solutions sur des temps plus longs.

Egalement, durant ce stage, j'ai travaillé avec des données issues d'une technologie nouvelle et en plein essor. J'ai donc pu non seulement mettre en application mais aussi approfondir les connaissances théoriques acquises au cours de ma formation sur la technologie LiDAR. J'ai aussi développé de nouvelles compétences dans des domaines variés tels que l'écologie ou les statistiques. Enfin, une grande rigueur méthodologique et un sens critique important par rapport aux résultats obtenus ont été nécessaires au bon déroulement de ce projet exploratoire. Je suis sûr que toutes ces compétences acquises pendant ce stage me seront fortement utiles dans mon travail futur.

BIBLIOGRAPHIE

- Amiri N, *Assessment of Marker-controlled Watershed Segmentation Algorithm for Individual Tree Top Detection and Crown Delineation*, University of Twente Faculty of Geo-Information and Earth Observation (ITC), 2014, 78 pages
- Baltsavias EP, *Airborne laser scanning: basic relations and formulas*, ISPRS Journal of Photogrammetry & Remote Sensing, 54, 1999, pp. 1999-214
- Binard C, *Introduction à la regression PLS*, 2012, 44p
- Bitterlich, W, *The relascope idea. Relative measurements in forestry*, SERBIULA (sistema Librum 2.0), 1984
- Donoghue D.N.M, Watt P.J, Cox N.J & Wilson J, *Remote sensing of species mixtures in conifer plantations using LiDAR height and intensity data*, Remote Sensing of Environment, 2007, 110, 509–5 F
- Dormann C, Elith J, Bacher S, Buchmann C, Carl G, Carré G, Diekötter T, García Márquez J, Gruber B, Lafourcade B, Leitão P, Münkemüller T, McClean C, Osborne P, Reineking B, Schröder B, Skidmore A, Zurell D, Lautenbach S, *Collinearity: A review of methods to deal with it and a simulation study evaluating their performance*, Ecography, 2013, 36. 27-46
- Dufort J, *Estimation automatisée de la hauteur des arbres à partir de données d'altimétrie laser*, Ecole de technologie supérieure, Université du Québec, 2000, 153p
- Gachet G, *Analyse et exploitation des données LiDAR aéroportés pour la caractérisation des milieux boisés de la Suisse*, Thèse de Doctorat de l'Ecole Polytechnique Fédérale de Lausanne, 2009, 278p
- Guisan A, Edwards AT, Hastie T, *Generalized linear and generalized additive models in studies of species distributions: setting the scene*, Ecological Modelling 157, 2002, 89-100
- Herrault PA, Larrieu L, Cordier S, Gimmi U, Lachat T, Ouin A, Sarthou JP, and Sheeren D, *Combined effects of area, connectivity, history and structural heterogeneity of woodlands on the species richness of hoverflies (Diptera : Syrphidae)*, Landscape Ecology, vol. 31, n°4, 2016, 877-893
- Hyypä J, Kelle O, Lehikoinen M, Inkinen M, *A segmentation-based method to retrieve stem volume estimates from 3D tree height models produced by laser scanners*, IEEE Trans. Geosci. Remote Sensing, 2001, 39, 969–975
- Larrieu L., Gonin P., *L'indice de biodiversité potentielle (ibp): une méthode simple et rapide pour évaluer la biodiversité potentielle des peuplements forestiers*, 2008

- Lefsky MA, Cohen WB, Acker SA, Spies TA, Parker GG, Harding D, *Lidar remote sensing of biophysical properties and canopy structure of forest Douglas-fir and western hemlock*, Remote Sensing of Environment, 1999b, 70: 339-361
- Lefsky MA, Cohen WB, Parker GG, Harding DJ, *Lidar remote sensing for ecosystem studies*, Bioscience; Jan 2002; 52, 1; Research Library, pg. 19
- Lesak L, Radeloff VC, Hawbaker TJ, Pidgeon AM, Gobakken T, Contrucci K, *Modeling forest songbird richness using LiDAR-derived forest structure*, Remote Sensing of Environment, 2011, 115(11):2823-2835
- Lesur D, *Rapport d'étude sur l'utilisation des données LiDAR pour la caractérisation des milieux forestiers de Mayotte*, 2011
- Parker, G. & Russ M.E., *The canopy surface and stand development: assessing forest canopy structure and complexity with near-surface altimetry*, Forest Ecology and Management, 2004, 189, 307–315
- Mallet C, *LIDAR aéroportés topographiques & bathymétriques*, cours à l'ENSG pour la filière IGAST, 2017
- Millennium Ecosystem Assessment, *Ecosystems and human well-being: synthesis*, Washington, DC: Island Press, 2015, ISBN 1-59726-040-1
- Morsdorf, F., Marell, A., Koetz, B., Cassagne, N., Pimont, F., Rigolot, E. & Allgöwer, B., *Discrimination of vegetation strata in a multi-layered Mediterranean forest ecosystem using height and intensity information derived from airborne laser scanning*, Remote Sensing of Environment, 2010, 114, 1403–1415
- Popescu S, Wynne R, Nelson R, *Measuring individual tree crown diameter with Lidar and assessing its influence on estimating forest volume and biomass*, Canadian Journal of Remote Sensing, 2003, 29. 564-577
- Rodriguez, M. A., Belmontes, J.A. & Hawkins, B.A. *Energy, water and large-scale patterns of reptile and amphibian species richness in Europe*. Acta Oecologica, 2005, 28, 65–70.
- Schwarzkopf, L. & Rylands, A.B. *Primate species richness in relation to habitat structure in amazonian rainforest fragments*, Biological Conservation, 1989, 48, 1–12.
- Simonson WD, Allen HD, Coomes DA, *Applications of airborne lidar for the assessment of animal species diversity*, Methods in Ecology and Evolution, 2014, 5, 719–729
- UICN France, *Panorama des services écologiques fournis par les milieux naturels en France – volume 1 : contexte et enjeux*. 2012, Paris, France
- Van Leeuwen M, Coops NC, Wulder MA, *Canopy surface reconstruction from a LiDAR point cloud using Hough transform*, Remote Sensing Letters 2010, 1, 125–132

Vierling K.T., Vierling L.A., Gould W.A., Martinuzzi S. & Clawges R.M, *Lidar: shedding new light on habitat characterization and modeling*, *Frontiers in Ecology and the Environment*, 2008, 6, 90–98

Wang L, Gong P, Biging GS, *Individual tree-crown delineation and treetop detection in high-spatial-resolution aerial imagery*, *Photogrammetry Engineering & Remote Sensing*, 2004, 70, 351–357

Weinacker H, Koch B, Heyder U, Weinacker R, *Development of filtering, segmentation and modelling modules for LiDAR and multispectral data as a fundamental of an automatic forest inventory system* *ISPRS Journal of Photogrammetry and Remote Sensing*, 2004, 36, 50–55

White JC, Wulder MA, Varhola A, Vastaranta M, Coops NC, Cook BD, PittD, Woods M, *Guide des meilleures pratiques pour générer des attributs d’inventaire forestier provenant de données obtenues par balayage laser aéroporté en utilisant une approche par zones*, *Ressources naturelles du Canada, Rapport d’information*, 2013, FI-X-010

Wing BM, Ritchie MW, Boston K, Cohen WB, Olsen MJ, *Individual snag detection using neighborhood attribute filtered airborne Lidar data*, *Remote Sensing of Environment*, 2015, 163: 165-179

Wulder MA, White JC, Nelson RF, Næsset E, Ørka HO, Coops NC, *Lidar sampling for large-area forest characterization: A review*, *Remote Sensing of Environment* 121, 2012, 196–209

Zhao K, Popescu S, Meng X, Pang Y, Agca M, *Characterizing forest canopy structure with Lidar composite metrics and machine learning*, *Remote Sensing of Environment* 115, 2011, Pages 1978-1996

Zhen Z, Quackenbush LJ, Stehman SV, Zhang L, *Agent-based region growing for individual tree crown delineation from airborne laser scanning (ALS) data*, *International Journal of Remote Sensing*, 2015, 36, 1965–1993

Zhen Z, Quackenbush LJ, Zhang L, *Trends in Automatic Individual Tree Crown Detection and Delineation - Evolution of LiDAR Data*, *Remote sensing*, 2016, 8, 333

Zuur AF, Ieno EN, Elphick CS, *A protocol for data exploration to avoid common statistical problems*, *Methods in Ecology and Evolution*, 2010, 1, 3–14

ANNEXES

Annexe 1 : Diagramme de GANTT

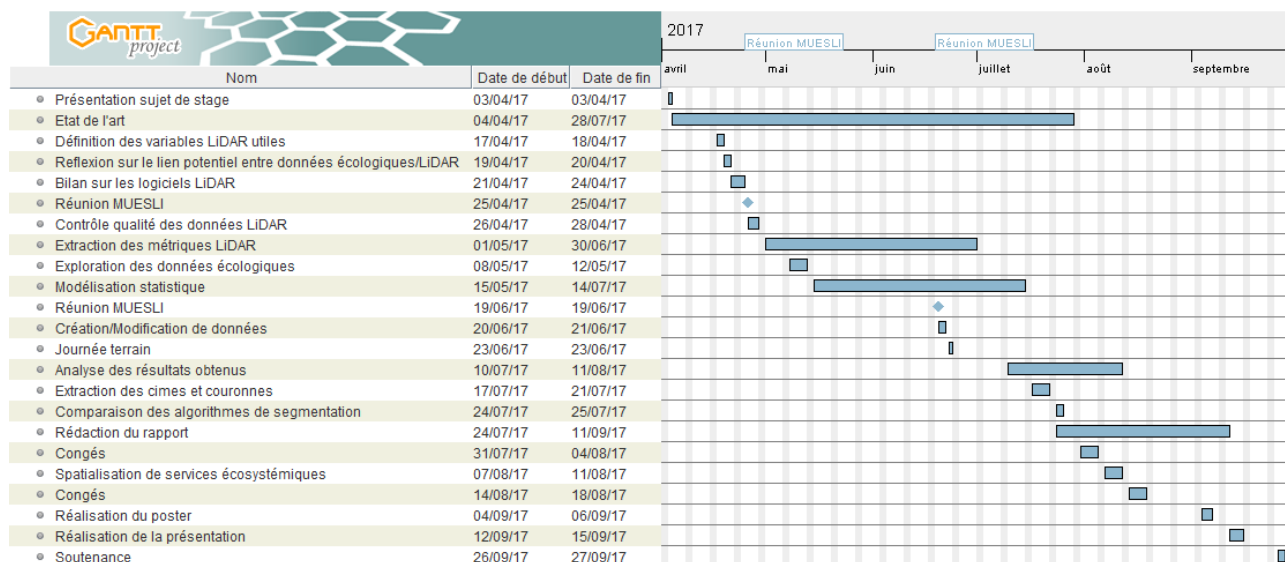


Figure 34 Diagramme de GANTT du stage.

Annexe 2 : Calcul de l'IBP

OBSERVATION DE DIX FACTEURS FAVORABLES A LA DIVERSITE DES ESPECES EN FORET
Fiche atlantique et continentale



FICHE DE SENSIBILISATION...

- pour les personnes peu familiarisées avec la gestion forestière, souhaitant découvrir les facteurs influençant la capacité d'accueil en espèces d'un peuplement forestier,
- ... **QUI NE PERMET PAS DE CALCULER L'INDICE DE BIODIVERSITE POTENTIELLE (IBP)**
- cette fiche est issue de l'IBP, mais avec des définitions simplifiées, en excluant les zones d'utilisation moins fréquentes (subalpin, stations peu fertiles), sans score ce qui limite la comparaison et le suivi des peuplements ;
- pour calculer l'IBP, utiliser la fiche téléchargeable sur internet : www.foretpriveefrancaise.com/ibp.

COMMENT OBSERVER LES 10 FACTEURS ?

- un relevé distinct par peuplement (zone homogène correspondant à une unité de gestion) ;
- parcourir suffisamment le peuplement pour en avoir une vue d'ensemble ;
- matériel nécessaire : ruban pour mesurer la grosseur des arbres, plans de la forêt, documentation (au verso).

ZONE D'UTILISATION



Références de mon peuplement (parcelle, lieu-dit, surface parcourue...) :

J'observe les 10 facteurs ci-dessous dans mon peuplement et je dénombre les éléments (en m'arrêtant de compter quand j'atteins le seuil sup.) Je coche la valeur trouvée sur le graphique de droite et j'en déduis la capacité d'accueil pour les espèces

7 facteurs liés au peuplement et à la gestion forestière		Combien de genres d'essences locales parmi cette liste ? Ne pas distinguer les espèces, par ex. présence d'Erable champêtre et d'Erable sycomore, ne compter qu'un Erable. Alisier, Cormier et Sorbier (= Sorbus) / Aulne / Bouleau / Charme / Charme houblon / Châtaignier / Chêne à feuilles caduques / Chêne à feuilles persistantes / Epicéa / Erable / Frêne / Hêtre / If / Mélèze / Merisier et Cerisier (=Prunus) / Noyer (commun) / Orme / Peuplier et Tremble / Pin / Poirier / Pommier / Sapin / Saule / Tilleul	Nombre d'essences 0 3 5
		Combien de strates présentes parmi les 4 strates suivantes : (un arbre peut être comptabilisé dans plusieurs strates) Ligneuse haute (20m) / Ligneuse intermédiaire (7m) / Ligneuse basse / Herbacées	Nombre de strates 0 3 4
		Combien de chandelles, d'arbres secs sur pied ou de souches hautes, de grosse circonférence ? 1,30 m du sol / gros bois morts : diamètre ≥ 40 cm (C ≥ 120 cm)	Nombre / ha 0 1 3
		Combien de troncs morts tombés au sol, de grosse circonférence ? 1 m de la base / gros bois morts : diamètre ≥ 40 cm (C ≥ 120 cm)	Nombre / ha 0 1 3
		Combien de très gros arbres vivants ? Très gros arbres : diamètre ≥ 70 cm (C ≥ 220 cm) / 1,3 m du sol	Nombre / ha 0 1 5
		Combien d'arbres vivants porteurs de microhabitats et combien de types de microhabitats différents ? Typologie présentée au verso.	Nbre d'arbres à microhabitats/ha 0 1 6 Nbre de types de microhabitats 0 1 3
		Quelle surface de milieux ouverts ? Un milieu ouvert est caractérisé par sa végétation particulière, avec des plantes à fleurs et des herbacées, et une floraison abondante au moins une partie de l'année. Il peut se présenter sous forme de trouées dans le peuplement ou de lisières (le long d'un chemin, en bordure de prairie...). Calcul du % de surface : $\frac{\text{somme des surfaces de milieux ouverts (long x larg)}}{\text{surface du peuplement (en m}^2\text{)}} \times 100$	ouverture 0% 1% 5% nulle faible élevée capacité d'accueil faible forte moy
3 facteurs liés au contexte		Ma forêt est-elle "ancienne", c'est-à-dire qu'elle n'a pas été défrichée dans le passé pour un usage agricole (au moins depuis 1850) ? En général, elle est alors mentionnée sur l'ancienne carte de l'État-major (1820-1866) (voir sur internet : www.geoportail.fr , rubrique Histoire).	non en partie oui capacité d'accueil faible forte
		Combien de types de milieux aquatiques ? Pas de surface minimale, typologie présentée au verso.	Nombre de types 0 1 2 capacité d'accueil faible forte
		Combien de types de milieux rocheux ? Surface minimale de 20 m ² ou longueur mini de 20 m pour les murettes en pierre sèche. Typologie présentée au verso.	Nombre de types 0 1 2 capacité d'accueil faible forte

Larrieu L. & Ganin P. - CNPF-IDF, INRA Dynafor - 10/01/2014

AIDE











- pour identifier les essences : utiliser la Flore forestière française (Editions IDF) ou Clés de forêt (www.onf.fr)
- pour estimer les nombres d'arbres à l'ha si vous ne connaissez pas la surface de vos parcelles : estimer la surface du peuplement en multipliant longueur par largeur moyenne ou faites une estimation moyenne sur 1 ha en visualisant un carré de 100 m x 100 m
- pour les typologies : voir ci-dessous ou définition détaillée sur internet : www.foretriveefrancaise.com/ibp (d'après IBP, version 2.7)
- pour toutes autres définitions : consulter les ouvrages sur la forêt (voir Editions IDF) ou internet (www.foretriveefrancaise.com)

La typologie des MICROHABITATS

<p>Cavité creusée par les Pics</p> 	<p>Cavité de pied à fond dur</p> 	<p>Plage de bois non carié* sans écorce</p> 	<p>Cavité évolutive* à terreau de tronc</p> 	<p>Cavité évolutive* à terreau de pied</p> 	<p>Cavité remplie d'eau : dendrotelmes</p> 
<p>Fente et écorce décollée</p> 	<p>Champignon polypores</p> 	<p>Coulée de sève active</p> 	<p>Charpentière* ou cime récemment brisée</p> 	<p>Bois mort dans le houppier*</p> 	<p>Liane (et gui)</p> 

* bois non carié : bois non protégé par l'écorce, peu décomposé ; cavité évolutive : présence de bois à un stade avancé de décomposition ; houppier : ensemble des ramifications vivantes de l'arbre (branches et rameaux) ; charpentière : grosse branche principale.

La typologie des MILIEUX AQUATIQUES

<p>Source (et suintement)</p> 	<p>Ruisseau, fossé humide non entretenu et petit canal</p> 	<p>Petit cours d'eau</p> 	<p>Rivière et fleuve (estuaire et delta)</p> 	<p>Bras mort</p> 
<p>Lac (et plan d'eau profond)</p> 	<p>Etang et lagune (et plan d'eau peu profond)</p> 	<p>Mare (et autres petits points d'eau)</p> 	<p>Tourbières</p> 	<p>Zone marécageuse</p> 

La typologie des MILIEUX ROCHEUX

<p>Falaise</p> 	<p>Dalle</p> 	<p>Lapiaz</p> 	<p>Grotte, gouffre</p> 	<p>Eboulis instable</p> 
<p>Amoncellement de blocs stables (blocs épars ou murettes)</p> 	<p>Affleurement de bancs de galets</p> 	<p>Chaos de blocs</p> 	<p>Rocher (paroi ou affleurement)</p> 	<p>Rocher (paroi ou affleurement)</p> 

COMMENT INTEGRER LES RESULTATS DU DIAGNOSTIC DANS LA GESTION ?

- chercher à conserver les éléments favorables et améliorer les autres, d'une manière compatible avec les objectifs de gestion et en tenant compte des autres facteurs (marché du bois, desserte, risques...)
- voir des exemples dans : "Diversité des espèces en forêt : pourquoi et comment l'intégrer dans la gestion. Emberger, Larrieu, Gonin, 2014"
- pour la biodiversité remarquable : faites un diagnostic complémentaire si besoin

* Fiche réalisée dans le cadre d'un programme national de développement de l'IBP mené par le Centre National de la Propriété Forestière (CNPF), avec la participation financière du Ministère de l'Ecologie, du Développement Durable et de l'Energie.
 * Remerciements aux relecteurs du CNPF, du Comité de pilotage du Programme IBP et aux propriétaires consultés, à Céline Emberger et Grégory Sajdak (CNPF-IDF) pour leur contribution à la maquette, ainsi qu'aux photographes Claude Berduco (C.B.), Mylène Préjant (M.P.) et Nicolas Goux (N.G.).
 * Référence : Larrieu L., Gonin P. : 2014 - Observation de dix facteurs favorables à la diversité des espèces en forêt. Fiche atlantique et continentale. CNPF-IDF, INRA Dynafor, 01/10/2014, 2 p. (en ligne : www.foretriveefrancaise.com/ibp)

Figure 35 Fiches de sensibilisation à l'IBP explicitant les facteurs entrant dans le calcul d'après Larrieu et Gonin.

Annexe 3 : Visualisation des données LiDAR

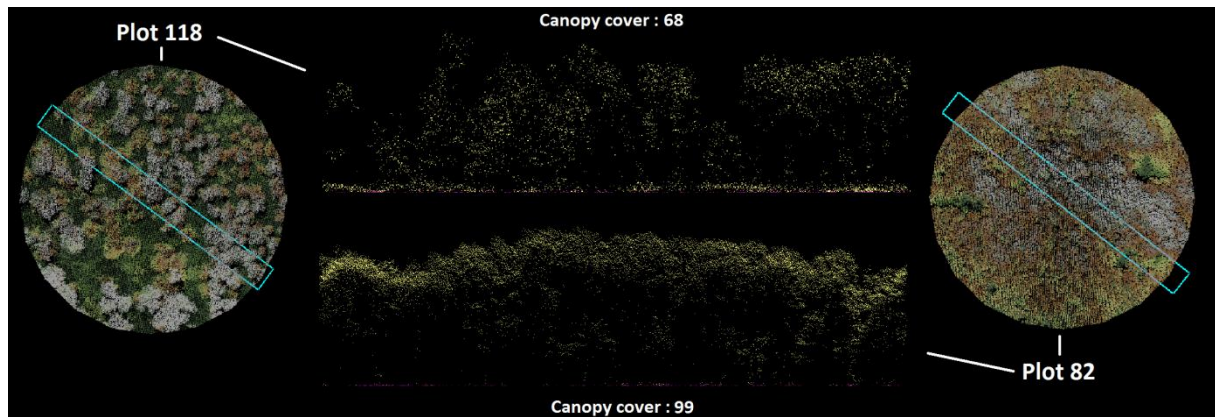


Figure 36 Visualisation des données LiDAR pour deux plots présentant des valeurs extrêmes de la couverture de canopée.

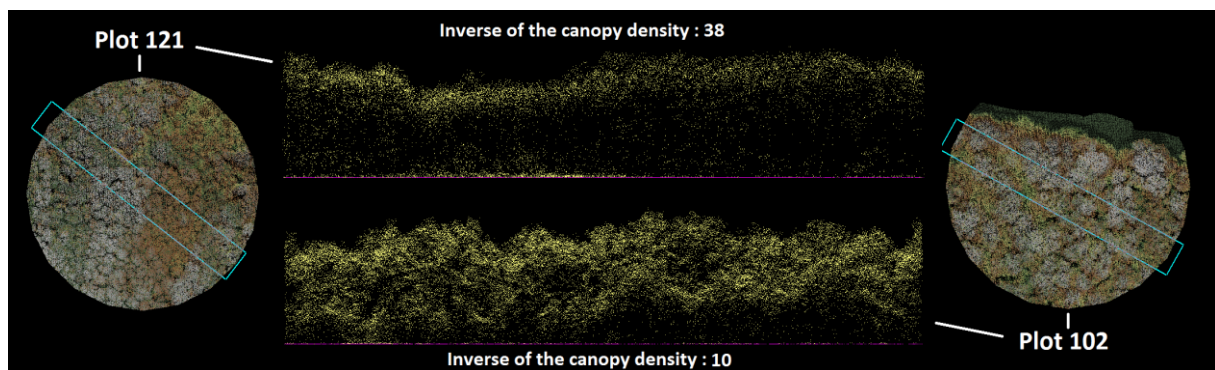


Figure 37 Visualisation des données LiDAR pour deux plots présentant des valeurs extrêmes de l'inverse de la densité de canopée.

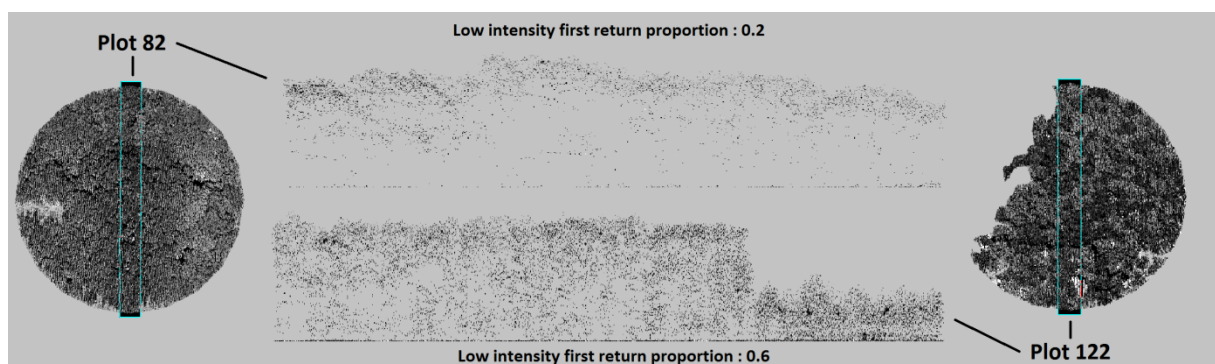


Figure 38 Visualisation des données LiDAR pour deux plots présentant des valeurs extrêmes de la proportion de 1ers retours de faible intensité.

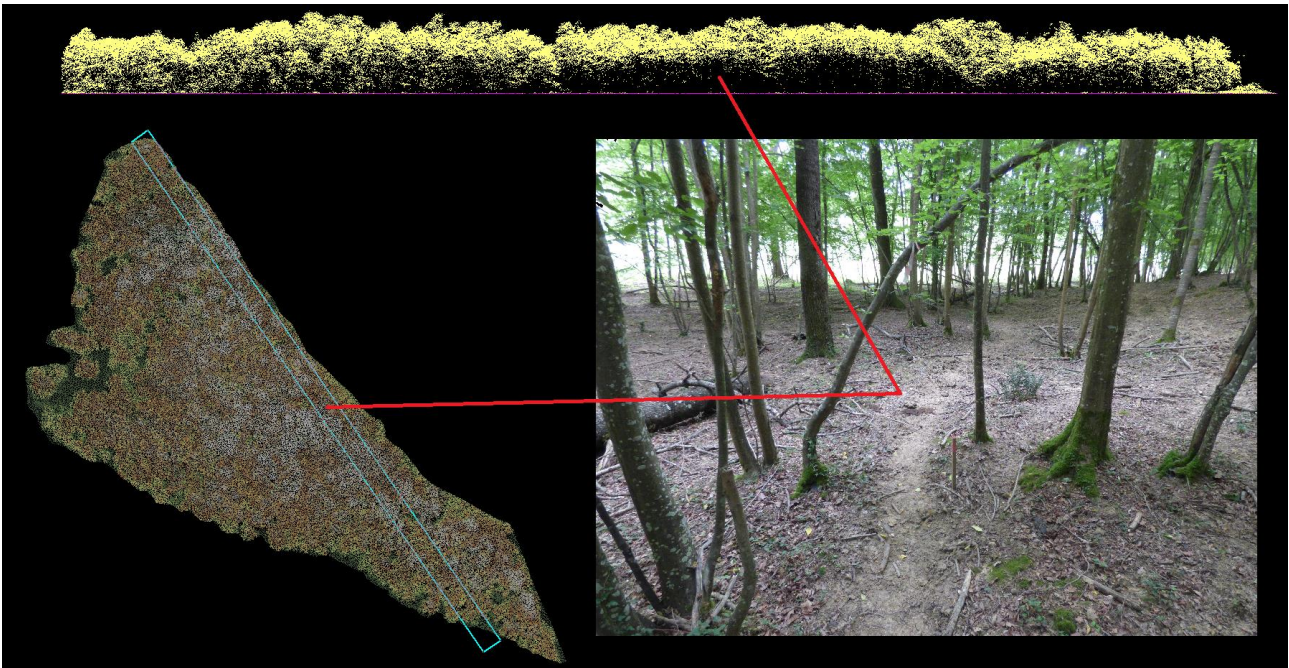
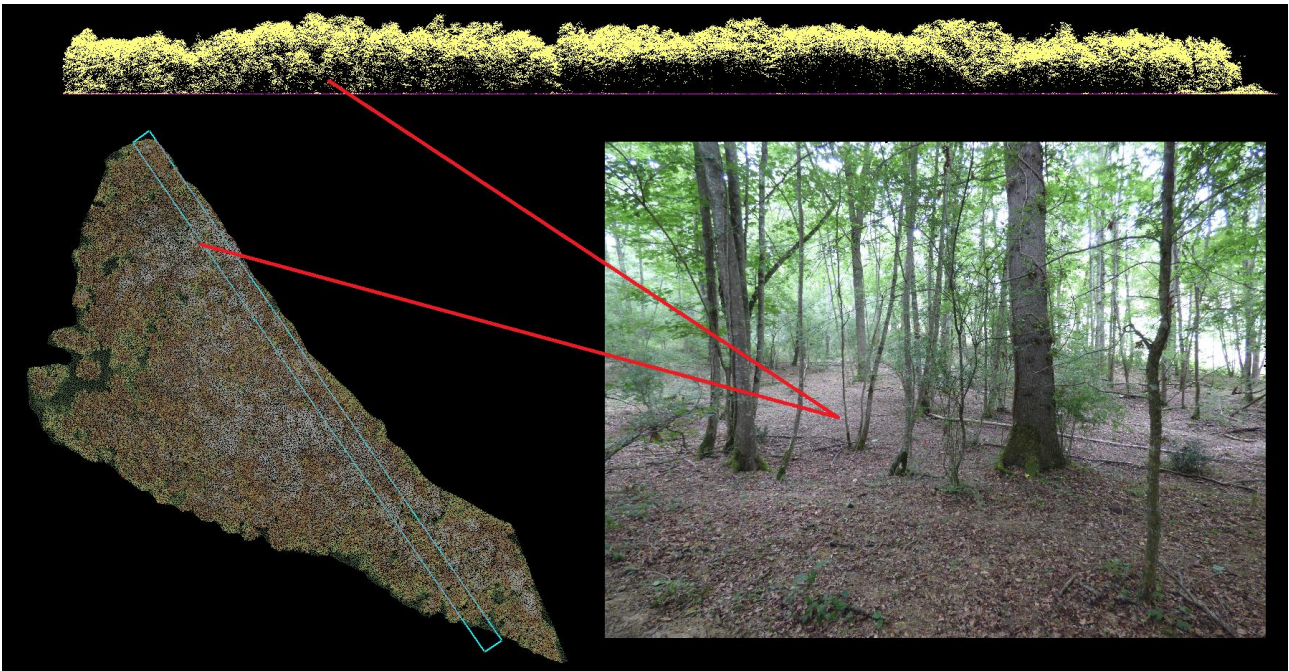


Figure 39 Visualisation terrain/nuage de points LiDAR pour deux zones d'un profil vertical.

Annexe 4 : Code d'extraction de variables LiDAR

a) Extraction à l'échelle des plots

```
echo off
mkdir %1\result
mkdir %1\tuilesLidar\newPlot

Fusion\PolyClipData /multifile /shape:1,* %1\ibp\placettesIBFVeget.shp %1\tuilesLidar\newPlot\clipTuiles.laz %1\tuilesLidar\*.laz

LAStools\bin\lasheight -i %1\tuilesLidar\newPlot\*.laz -replace_z -odix _height -olaz

LAStools\bin\lasindex -i %1\tuilesLidar\newPlot\*_height.laz

LAStools\bin\lascanopy -i %1\tuilesLidar\newPlot\*_height.laz ^
    -merged ^
    -cover_cutoff 3.0 ^
    -cov -dns -gap ^
    -height_cutoff 2.0 ^
    -centroids ^
    -c 2.0 5.0 10.0 50.0 ^
    -lcp %1\ibp\placettesIBFVeget.shp ^
    -min -max -avg -qav -std -ske -kur ^
    -p 1 5 10 25 50 75 90 95 99 ^
    -b 30 50 80 90 ^
    -d 2.0 5.0 10.0 50.0 ^
    -int_min -int_max -int_avg -int_qav -int_std -int_ske -int_kur ^
    -int_p 1 5 10 25 50 75 90 95 99 ^
    -int_c 0 128 256 1024 ^
    -int_d 0 128 256 1024 ^
    -o %1\result\L_MetricsIBP.csv

Fusion\CloudMetrics /above:3.0 /new /minht:2.0 /id %1\tuilesLidar\newPlot\*_height.laz %1\result\F_MetricsIBP.csv
```

Figure 40 Script de commandes batch permettant de lancer l'extraction des variables LiDAR avec LAStools et FUSION à l'échelle de plots.

b) Extraction par grille

```
echo off
mkdir result
mkdir result\tuile
mkdir result\csv_faibleInt
mkdir result\csv_allPoints
mkdir result\csv
mkdir result_csv_bois
C:\Users\sig\software\Fusion\PolyClipData /multifile /shape:6,* shape\veget_zone_lidar_fusion_split_areal.shp result\tuile\clipVeget.laz tuiles1\*.laz
:: intensite faible
C:\Users\sig\software\LAStools\bin\las2las -i result\tuile\clipVeget_*.laz -drop_intensity_above 60 -odix_intInf60 -olaz
C:\Users\sig\software\Fusion\CloudMetrics /above:3.0 /new /minht:2.0 /id result\tuile\*_intInf60.laz result\csv\MetricsBoisFInt.csv

for /f "tokens=1,31* delims=," %%a in (result\csv\MetricsBoisFInt.csv) do (
    for /f "tokens=31* delims=," %%A in ("%%b") do (
        for /f "tokens=23 delims=," %%B in ("%%A") do (
            echo %%a,%%B >> result\csv_faibleInt\gridMetricsFInt.csv
        )
    )
)
del result\tuile\*_intInf60.laz /s /q

C:\Users\sig\software\LAStools\bin\lasheight -i result\tuile\clipVeget_*.laz -replace_z -odix_height -olaz
C:\Users\sig\software\LAStools\bin\lasindex -i result\tuile\*_height.laz
C:\Users\sig\Fusion\CloudMetrics /above:3.0 /new /minht:2.0 /id result\tuile\*_height.laz result\csv\MetricsBois.csv

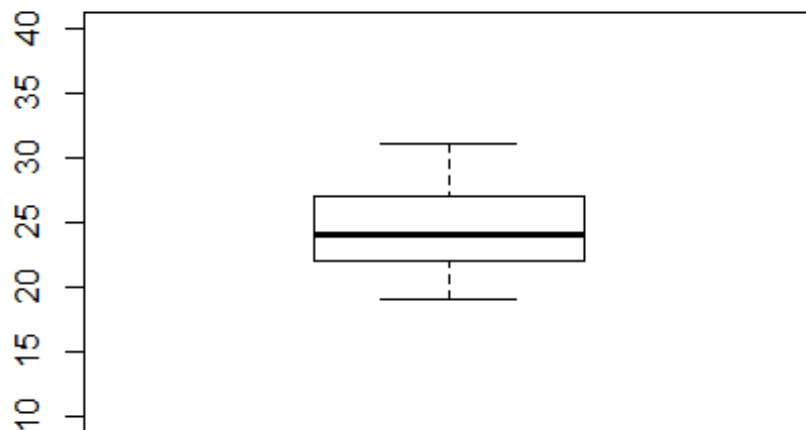
for /f "tokens=1,17,18,31* delims=," %%a in (result\csv\MetricsBois.csv) do (
    for /f "tokens=31* delims=," %%A in ("%%d") do (
        for /f "tokens=23,24 delims=," %%B in ("%%A") do (
            echo %%a,%%b,%%c,%%B,%%C >> result\csv_allPoints\gridMetrics.csv
        )
    )
)
rmdir result\tuile /s /q
```

Figure 41 Script de commandes batch permettant de lancer l'extraction des variables LiDAR avec LAStools et FUSION par grille.

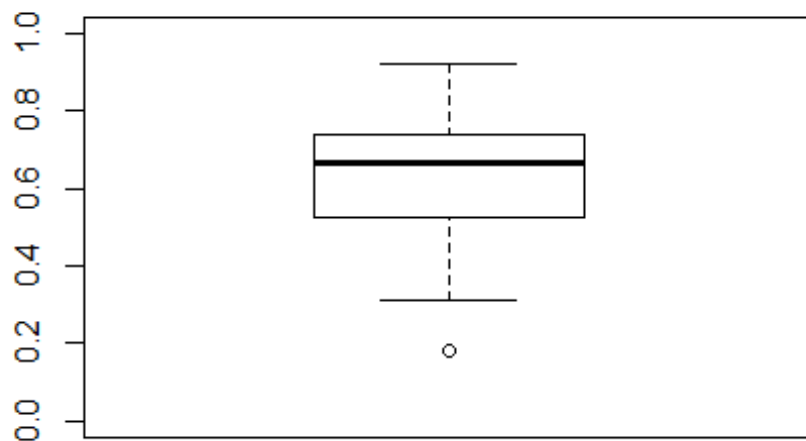
Annexe 5 : Résultats de la modélisation statistique par GLM

Boîtes à moustaches des variables écologiques à l'échelle des bois :

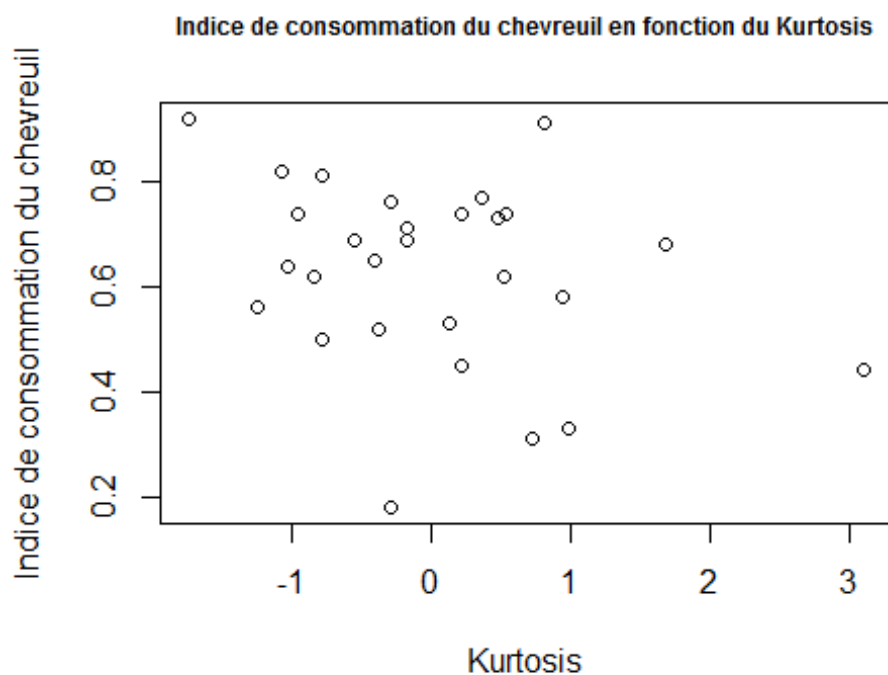
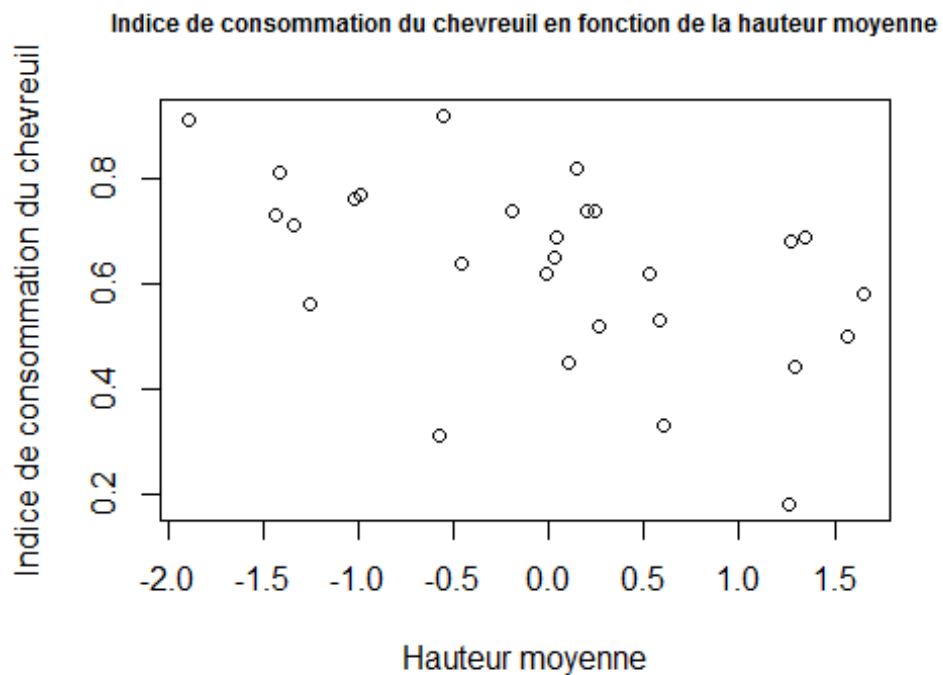
Boxplot de l'IBP



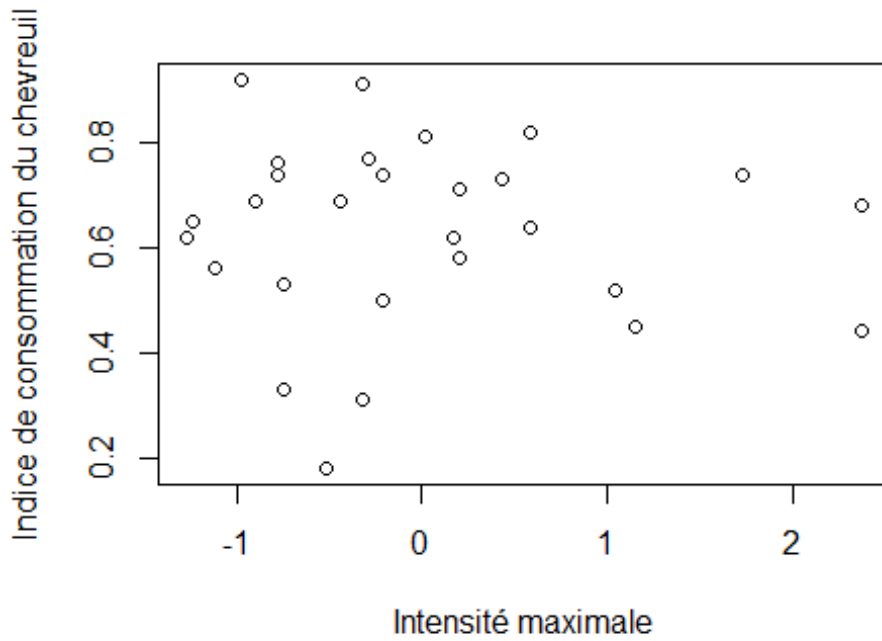
Boxplot de l'indice de consommation du chevreuil



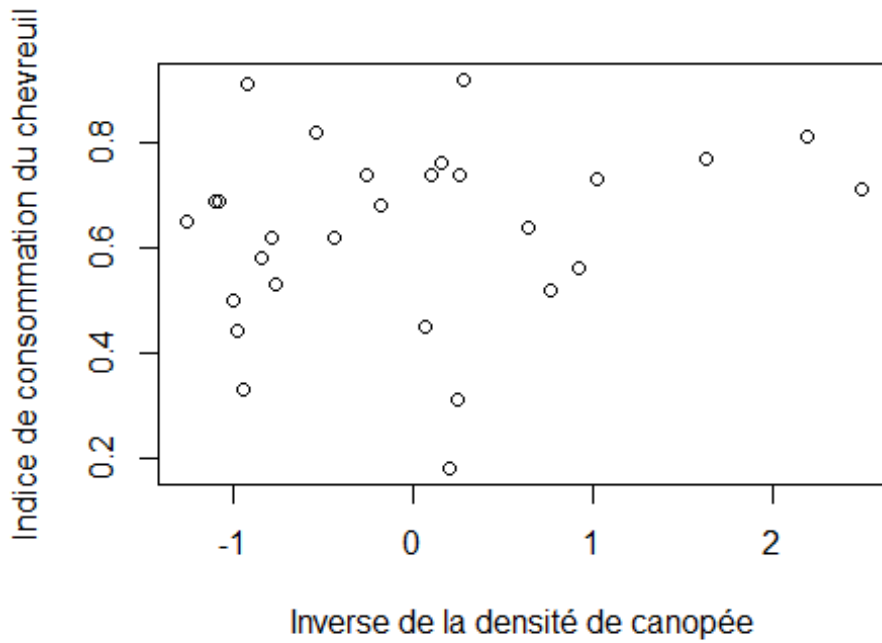
Diagrammes de dispersion de l'indice de consommation du chevreuil en fonction des variables LiDAR non corrélées à l'échelle des bois :



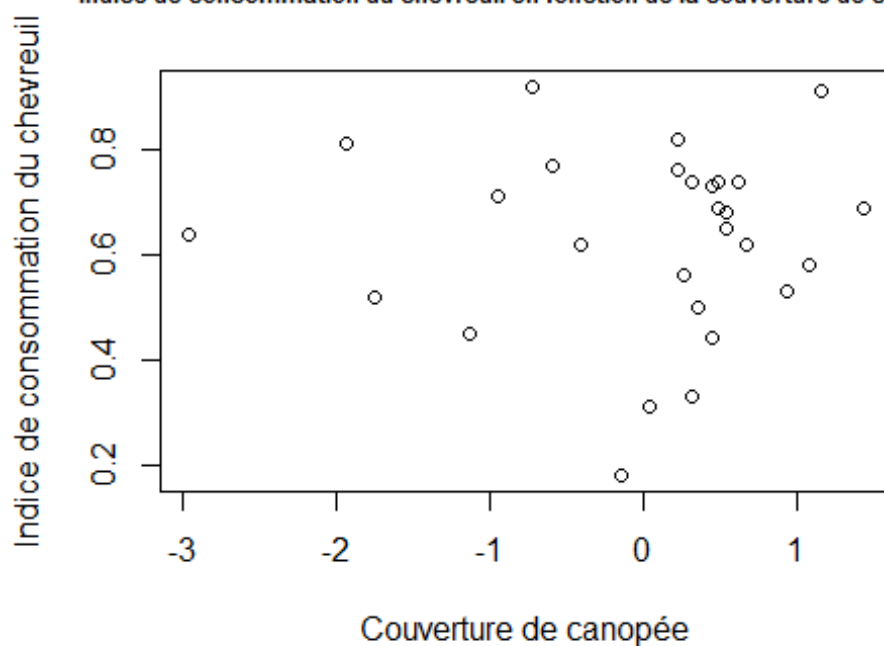
Indice de consommation du chevreuil en fonction de l'intensité maximale



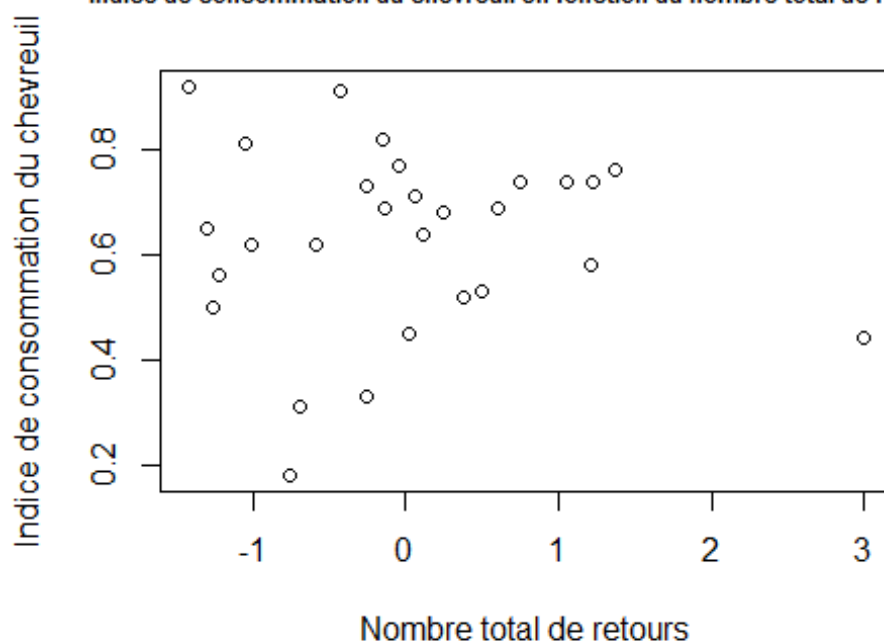
Indice de consommation du chevreuil en fonction de l'inverse de la densité de canopée



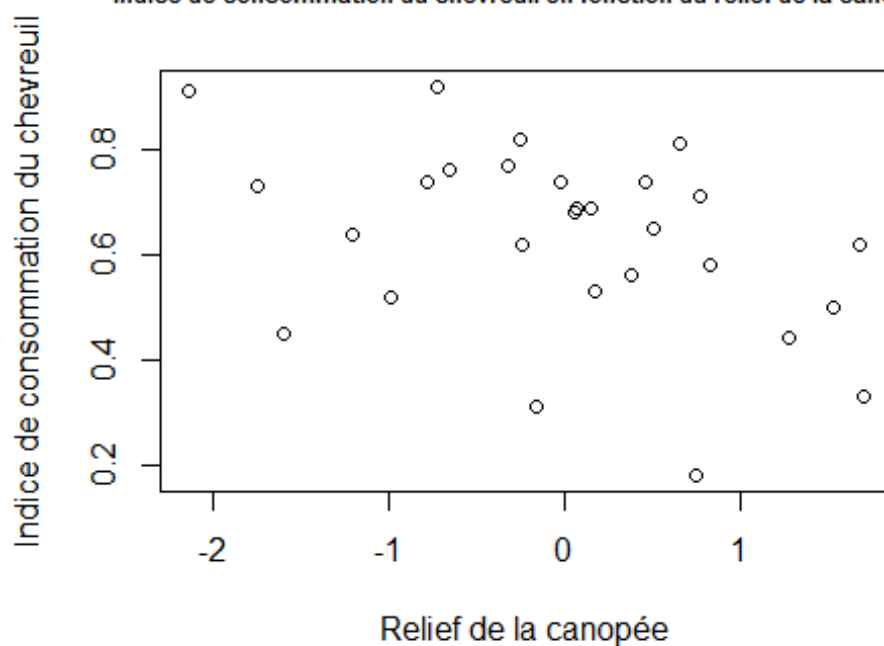
Indice de consommation du chevreuil en fonction de la couverture de canopée



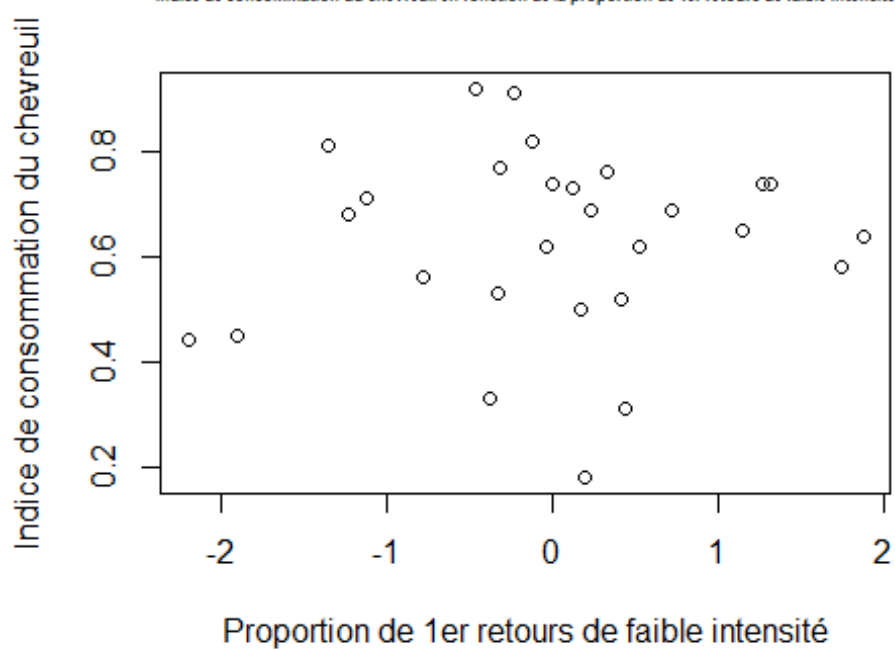
Indice de consommation du chevreuil en fonction du nombre total de retours

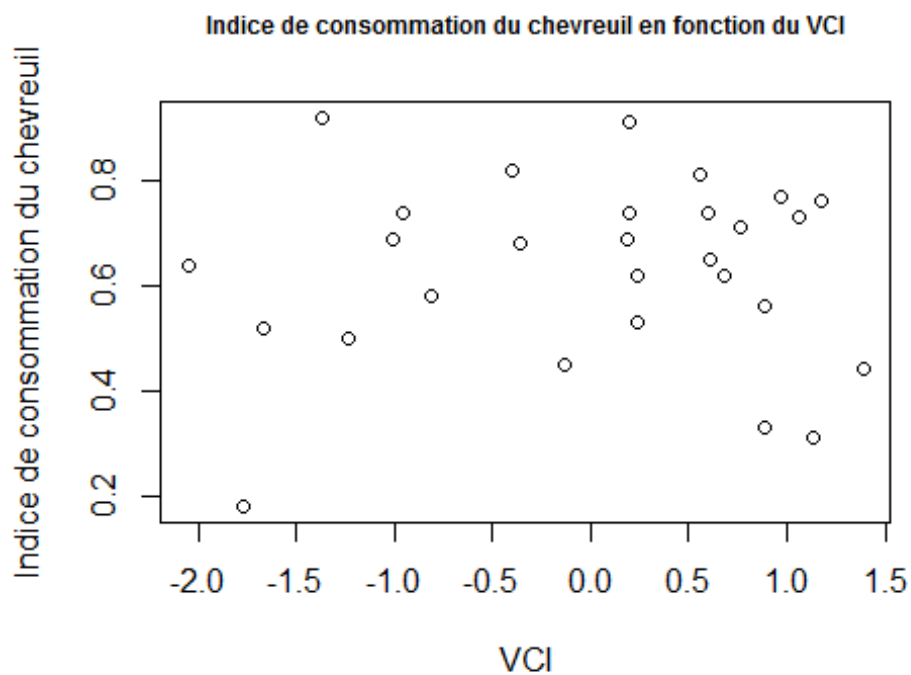


Indice de consommation du chevreuil en fonction du relief de la canopée

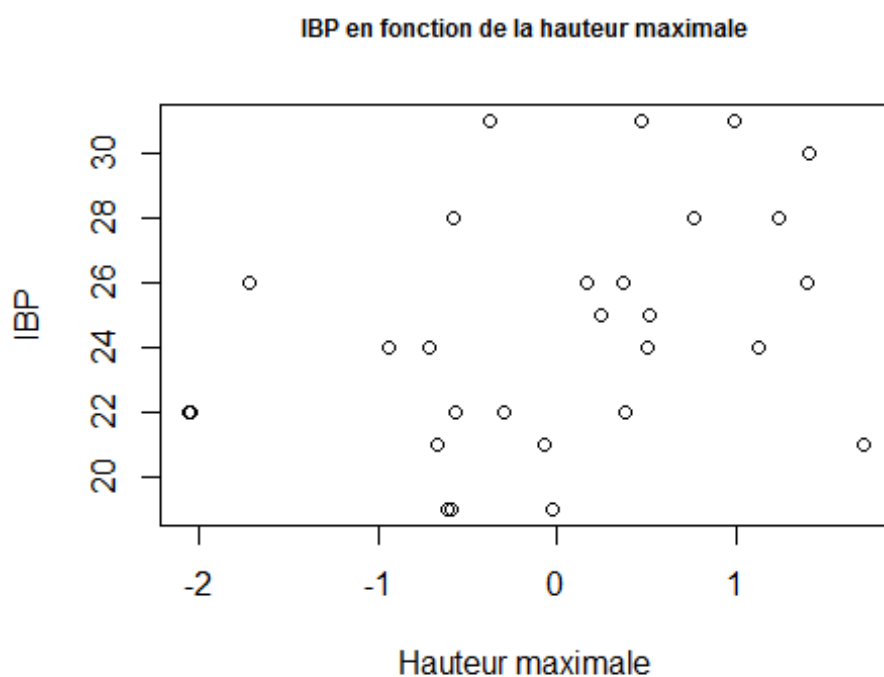


Indice de consommation du chevreuil en fonction de la proportion de 1er retours de faible intensité

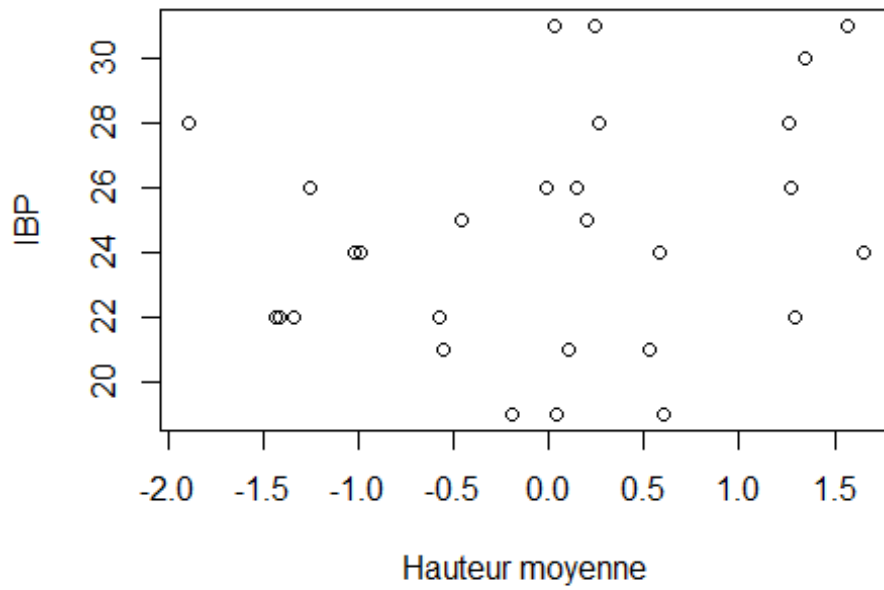




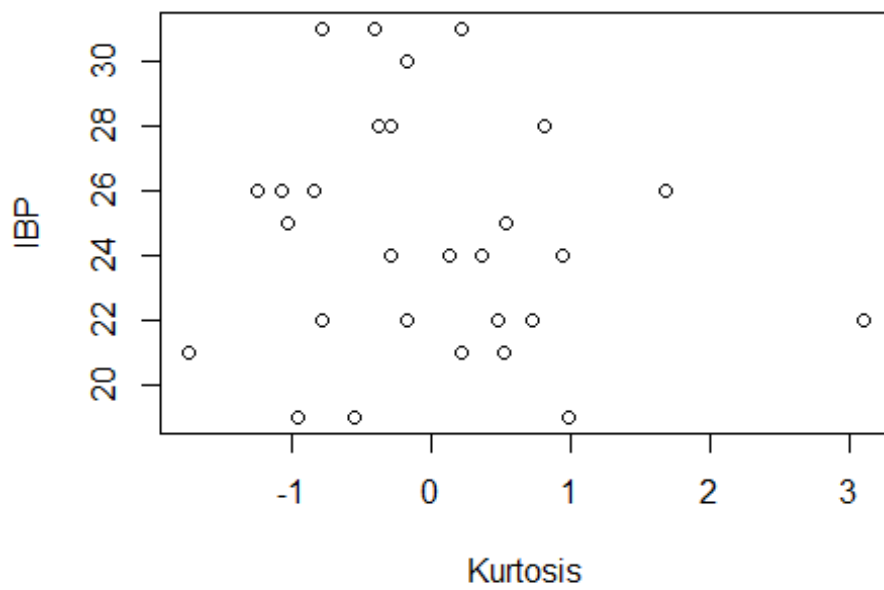
Diagrammes de dispersion de l'IBP en fonction des variables LiDAR non corrélées à l'échelle des bois :



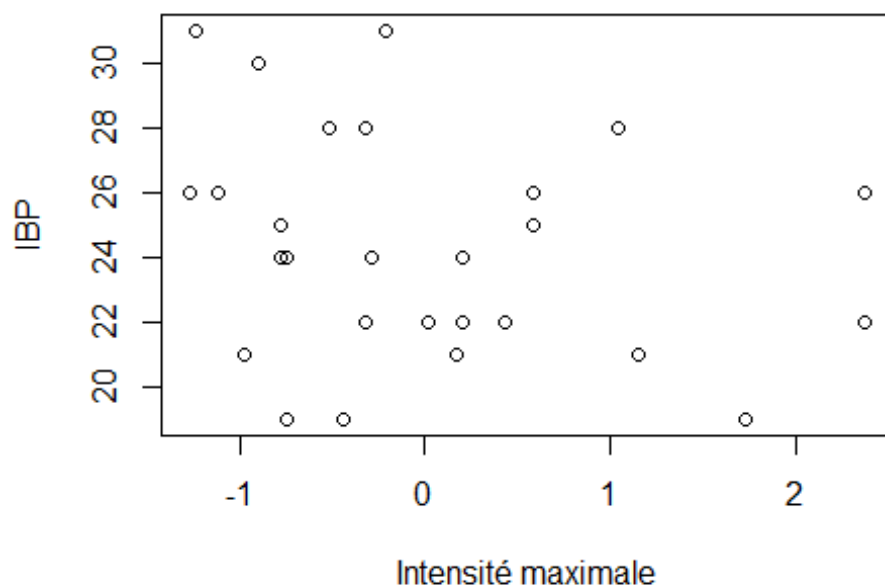
IBP en fonction de la hauteur moyenne



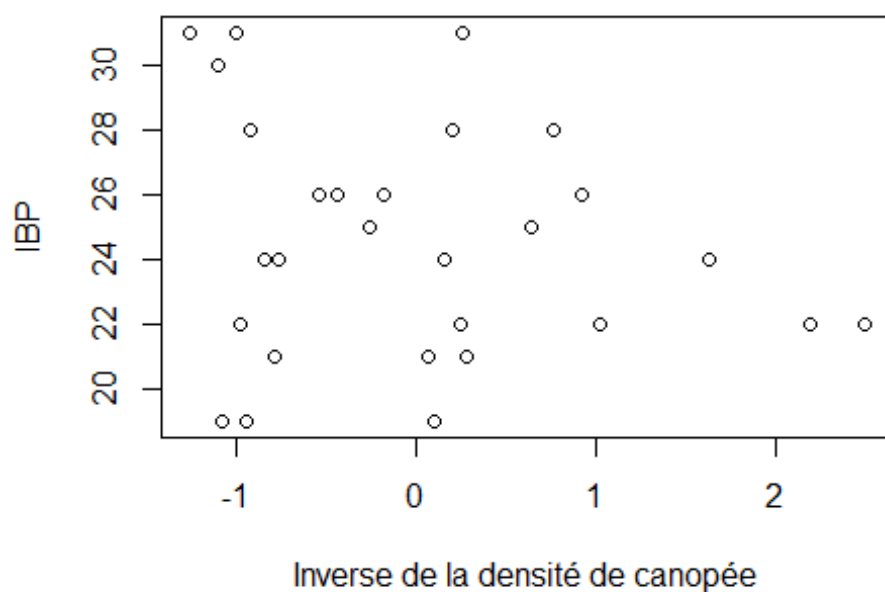
IBP en fonction du Kurtosis



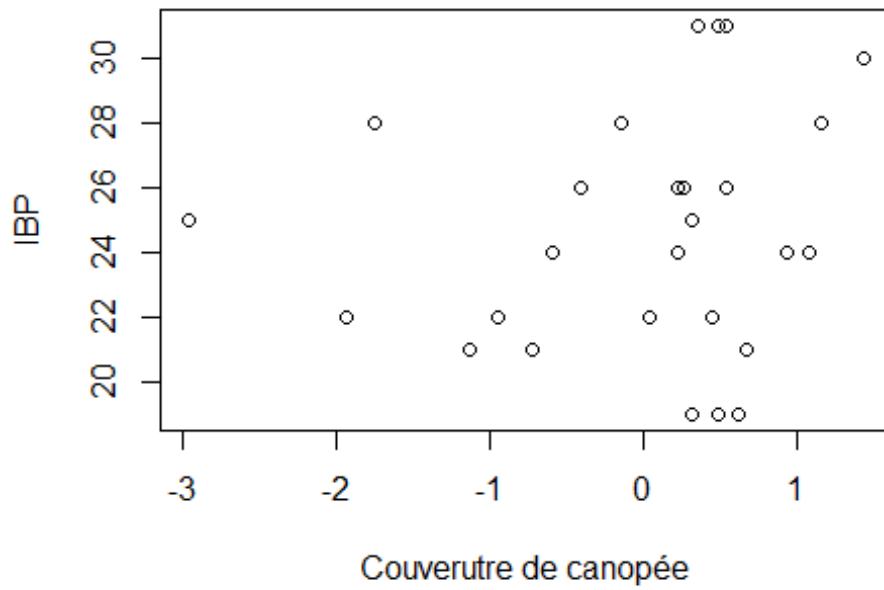
IBP en fonction de l'intensité maximale



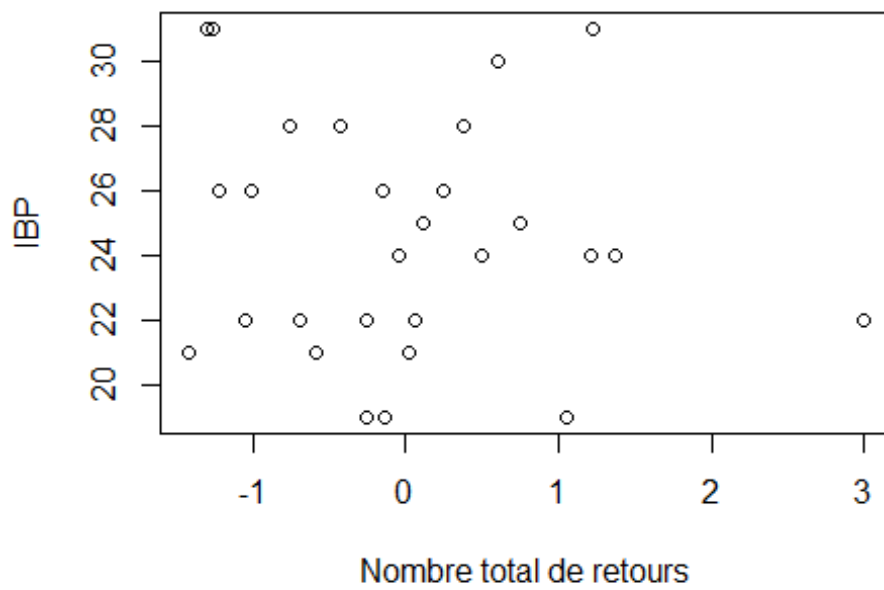
IBP en fonction de l'inverse de la densité de canopée



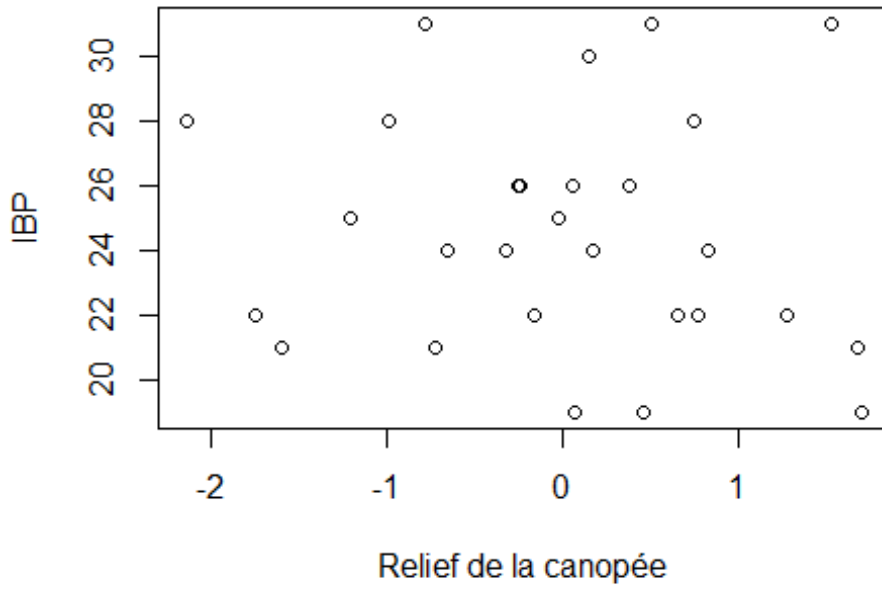
IBP en fonction de la couverture de canopée



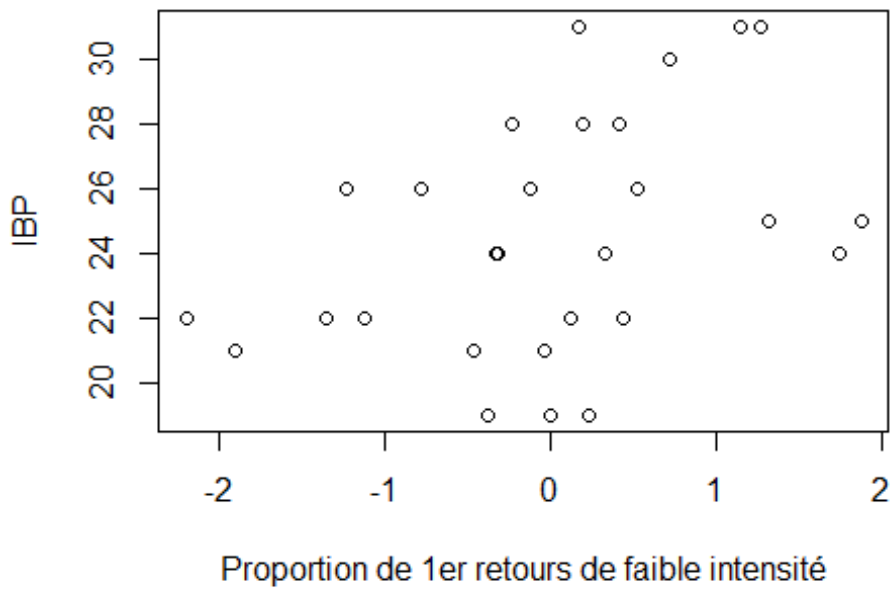
IBP en fonction du nombre total de retours

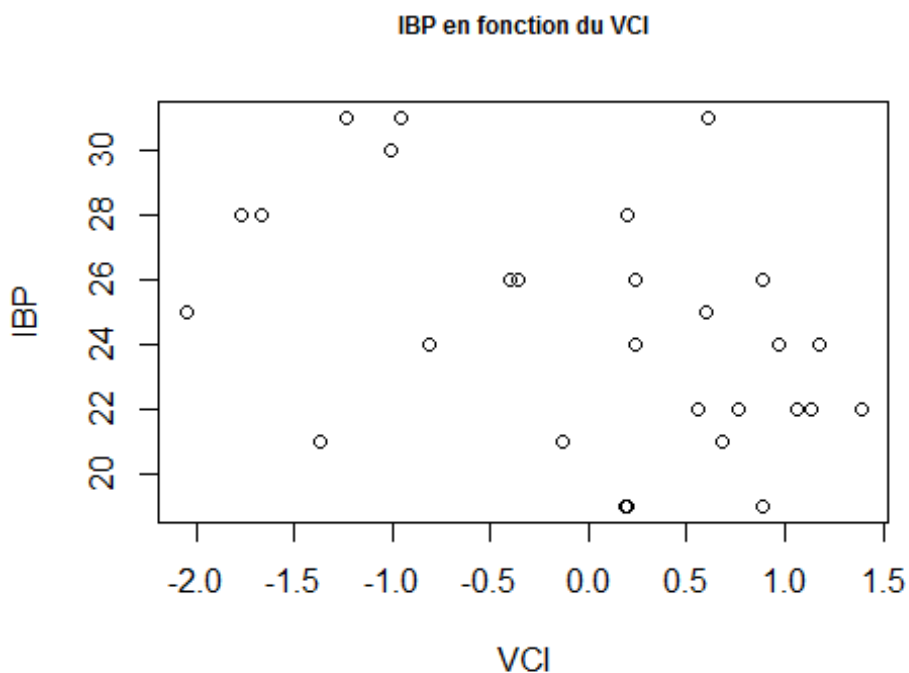


IBP en fonction du relief de la canopée



IBP en fonction de la proportion de 1er retours de faible intensité





Résultats de la modélisation statistique par GLM modèle n°15

```
# Selection du meilleur modèle (AIC Le plus faible)
mod_select = glm(formula = as.numeric(dataBoisUseful$ic_global) ~ cut2_cov +
  F_Elev.maximum + F_Elev.kurtosis + F_Total.all.returns +
  F_Canopy.relief.ratio, family = "gaussian", data =
dataMetricLidScale)
summary(mod_select)

##
## Call:
## glm(formula = as.numeric(dataBoisUseful$ic_global) ~ cut2_cov +
##     F_Elev.maximum + F_Elev.kurtosis + F_Total.all.returns +
##     F_Canopy.relief.ratio, family = "gaussian", data = dataMetricLidScale)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.30081 -0.07517  0.01807  0.07975  0.24002
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.63000    0.02778  22.675 < 2e-16 ***
## cut2_cov       0.04854    0.03127   1.552  0.13489
## F_Elev.maximum -0.07315    0.03113  -2.350  0.02817 *
## F_Elev.kurtosis -0.07291    0.03713  -1.964  0.06231 .
## F_Total.all.returns  0.05505    0.03629   1.517  0.14345
## F_Canopy.relief.ratio -0.08881    0.03055  -2.907  0.00818 **
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.02161529)
##
## Null deviance: 0.84040  on 27  degrees of freedom
## Residual deviance: 0.47554  on 22  degrees of freedom
## AIC: -20.654
##
## Number of Fisher Scoring iterations: 2
# plot(mod_select)
# Analyse de la deviance du modele
D2 <- (mod_select$null.deviance - mod_select$deviance) /
mod_select$null.deviance
D2
## [1] 0.4341548
```

Résultats de la modélisation statistique par GLM modèle n°8

```
# Selection du meilleur modele (AIC le plus faible)
mod_select = glm(formula = cbind(dataChenilleUseful$pred,
dataChenilleUseful$tot_chenilles - dataChenilleUseful$pred) ~
F_Elev.mean + F_Elev.kurtosis + F_Total.all.returns +
F_Canopy.relief.ratio +
myTableScale$propFirst + myTableScale$VCI +
as.numeric(dataBoisUseful$ab_tot_ois),
family = "binomial", data = dataMetricLidScale)
summary(mod_select)
##
## Call:
## glm(formula = cbind(dataChenilleUseful$pred,
dataChenilleUseful$tot_chenilles -
## dataChenilleUseful$pred) ~ F_Elev.mean + F_Elev.kurtosis +
## F_Total.all.returns + F_Canopy.relief.ratio + myTableScale$propFirst +
## myTableScale$VCI + as.numeric(dataBoisUseful$ab_tot_ois),
## family = "binomial", data = dataMetricLidScale)
##
## Deviance Residuals:
## Min 1Q Median 3Q Max
## -2.5387 -1.4580 0.0160 0.7197 3.1658
##
## Coefficients:
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.27421 1.11883 -0.245 0.806390
## F_Elev.mean -1.08153 0.19253 -5.617 1.94e-08
## F_Elev.kurtosis 0.37090 0.16267 2.280 0.022600
## F_Total.all.returns 0.41786 0.15853 2.636 0.008391
## F_Canopy.relief.ratio 0.57710 0.15728 3.669 0.000243
## myTableScale$propFirst -0.36878 0.14207 -2.596 0.009437
```

```

## myTableScale$VCI                -1.01968    0.22462  -4.540  5.63e-06
## as.numeric(dataBoisUseful$ab_tot_ois) -0.09631    0.06106  -1.577  0.114748
##
## (Intercept)
## F_Elev.mean                      ***
## F_Elev.kurtosis                   *
## F_Total.all.returns               **
## F_Canopy.relief.ratio             ***
## myTableScale$propFirst            **
## myTableScale$VCI                  ***
## as.numeric(dataBoisUseful$ab_tot_ois)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 110.807  on 27  degrees of freedom
## Residual deviance:  66.594  on 20  degrees of freedom
## AIC: 157.26
##
## Number of Fisher Scoring iterations: 5

# plot(mod_select)

# Analyse de La deviance du modele
D2 <- (mod_select$null.deviance - mod_select$deviance) /
mod_select$null.deviance
D2
## [1] 0.3990087

```

Résultats de la modélisation statistique par GLM modèle n°9

```

# Selection du meilleur modele (AIC Le plus faible)
mod_select = glm(formula = cbind(dataChenilleUseful$pred,
dataChenilleUseful$tot_chenilles - dataChenilleUseful$pred) ~
                F_Elev.mean + as.numeric(dataNDVIUseful$NDVI_mean) +
                F_Elev.kurtosis + F_Int.maximum + F_Canopy.relief.ratio +
                myTableScale$VCI, family = "binomial", data =
dataMetricLidScale)
summary(mod_select)

##
## Call:
## glm(formula = cbind(dataChenilleUseful$pred,
dataChenilleUseful$tot_chenilles -
##      dataChenilleUseful$pred) ~ F_Elev.mean +
as.numeric(dataNDVIUseful$NDVI_mean) +
##      F_Elev.kurtosis + F_Int.maximum + F_Canopy.relief.ratio +
##      myTableScale$VCI, family = "binomial", data = dataMetricLidScale)

```



```

##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.6507  -0.9844  -0.2641   0.6828   3.3637
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)         4.7667     2.8211   1.690 0.091097
## F_Elev.mean        -1.0888     0.1907  -5.709 1.14e-08
## as.numeric(dataNDVIUseful$NDVI_mean) -7.5870     3.1535  -2.406 0.016132
## F_Elev.kurtosis     0.4478     0.1592   2.813 0.004908
## F_Int.maximum       0.3190     0.1181   2.702 0.006895
## F_Canopy.relief.ratio 0.5033     0.1502   3.351 0.000806
## myTableScale$VCI    -0.7016     0.1854  -3.785 0.000154
##
## (Intercept)          .
## F_Elev.mean          ***
## as.numeric(dataNDVIUseful$NDVI_mean) *
## F_Elev.kurtosis      **
## F_Int.maximum        **
## F_Canopy.relief.ratio ***
## myTableScale$VCI    ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 110.807  on 27  degrees of freedom
## Residual deviance:  63.745  on 21  degrees of freedom
## AIC: 152.42
##
## Number of Fisher Scoring iterations: 5

# plot(mod_select)

# Analyse de la deviance du modele
D2 <- (mod_select$null.deviance - mod_select$deviance) /
mod_select$null.deviance
D2
## [1] 0.4247198

```

Résultats de la modélisation statistique par GLM modèle n°1

```

# Selection du meilleur modele (AIC le plus faible)
mod_select = glm(formula = dataIBPUseful$ibp ~ cut2_cov + F_Int.maximum +
myTableScale$VCI, family = "gaussian",
                 data = dataMetricLidUseful)
summary(mod_select)

##
## Call:

```

```

## glm(formula = dataIBPUseful$ibp ~ cut2_cov + F_Int.maximum +
##     myTableScale$VCI, family = "gaussian", data = dataMetricLidUseful)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -5.9940  -1.8176   0.3556   1.7992   6.6053
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -20.724246  31.317351  -0.662   0.5144
## cut2_cov      0.445337   0.298389   1.492   0.1486
## F_Int.maximum 0.009123   0.026488   0.344   0.7335
## myTableScale$VCI -1.399053  0.699910  -1.999   0.0571 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 11.02981)
##
## Null deviance: 362.96  on 27  degrees of freedom
## Residual deviance: 264.72  on 24  degrees of freedom
## AIC: 152.36
##
## Number of Fisher Scoring iterations: 2
# Analyse de La deviance du modele
D2 <- (mod_select>null.deviance - mod_select$deviance) /
mod_select>null.deviance
D2
## [1] 0.2706848

```

Résultats de la modélisation statistique par GLM modèle n°2

Selection du meilleur modèle (AIC Le plus faible)

```

mod_select = glm(formula = dataIBPUseful$ibp ~ cut2_cov + F_Elev.kurtosis +
F_Int.maximum + myTableScale$VCI + as.numeric(dataNDVIUseful$NDVI_var), family =
"gaussian", data = dataMetricLidUseful)

```

```
summary(mod_select)
```

Call:

```

glm(formula = dataIBPUseful$ibp ~ cut2_cov + F_Elev.kurtosis +
      F_Int.maximum + myTableScale$VCI + as.numeric(dataNDVIUseful$NDVI_var),
      family = "gaussian", data = dataMetricLidUseful)

```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-----	----	--------	----	-----

-6.8224 -1.9929 0.5358 2.0591 5.7239

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.07165	34.92517	-0.031	0.9758
cut2_cov	0.20250	0.35078	0.577	0.5696
F_Elev.kurtosis	4.46773	3.50775	1.274	0.2161
F_Int.maximum	-0.01370	0.03254	-0.421	0.6779
myTableScale\$VCI	-1.74763	0.74470	-2.347	0.0283 *
as.numeric(dataNDVIUseful\$NDVI_var)	344.21227	417.59512	0.824	0.4186

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 10.99059)

Null deviance: 362.96 on 27 degrees of freedom

Residual deviance: 241.79 on 22 degrees of freedom

AIC: 153.83

Number of Fisher Scoring iterations: 2

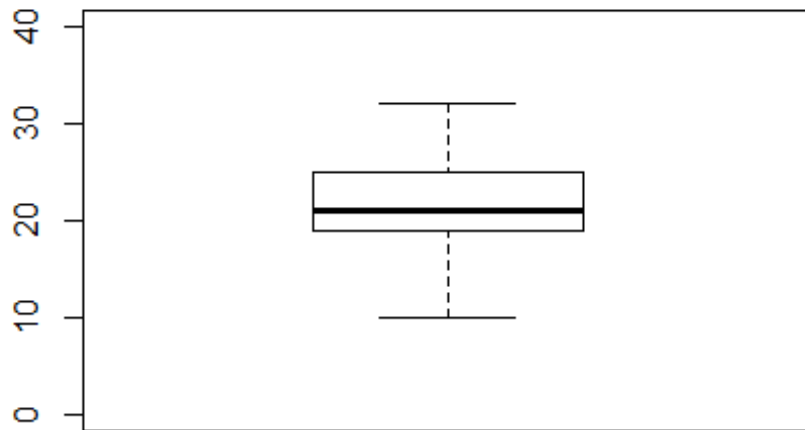
```
# Analyse de La deviance du modele > D2 <- (mod_select$null.deviance -  
mod_select$deviance) / mod_select$null.deviance
```

D2

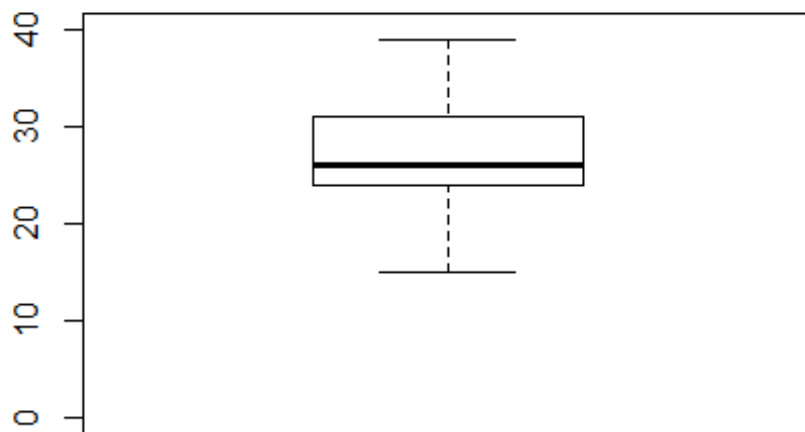
[1] 0.3338381

Boîtes à moustaches de l'iBP (total et partie gestion données 2011) à l'échelle des plots de 1ha :

Boxplot de l'iBP partie gestion

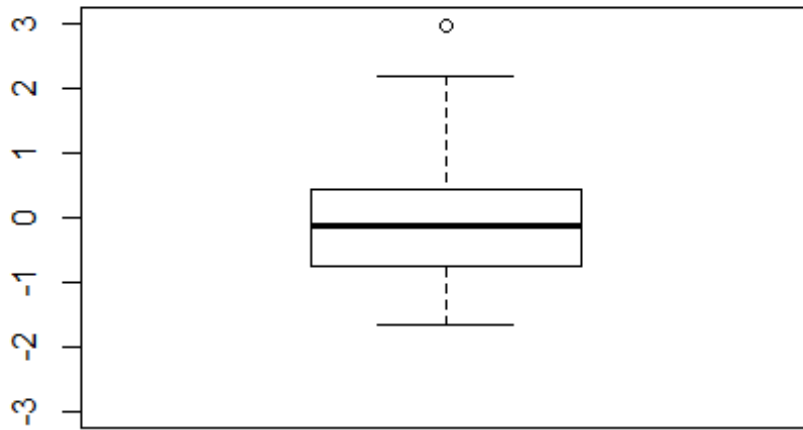


Boxplot de l'iBP

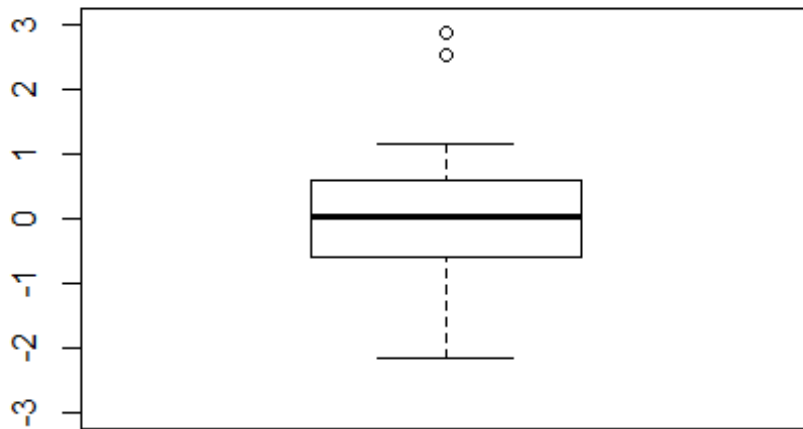


Boîtes à moustaches des variables LiDAR non corrélées à l'échelle des plots de 1ha :

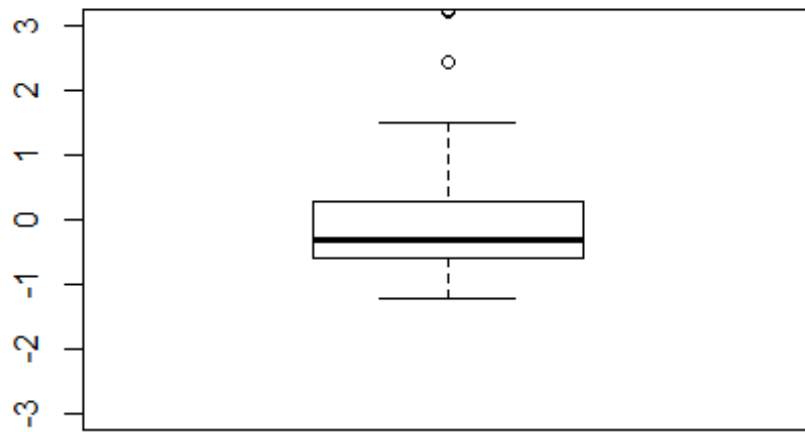
Boxplot de la hauteur maximale



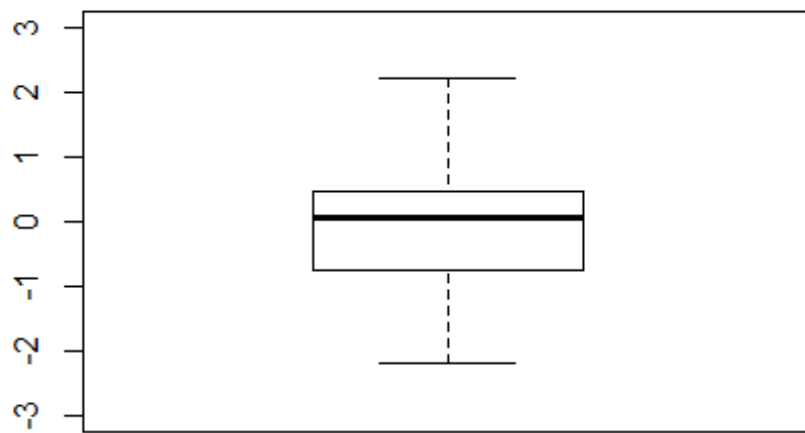
Boxplot de la hauteur moyenne



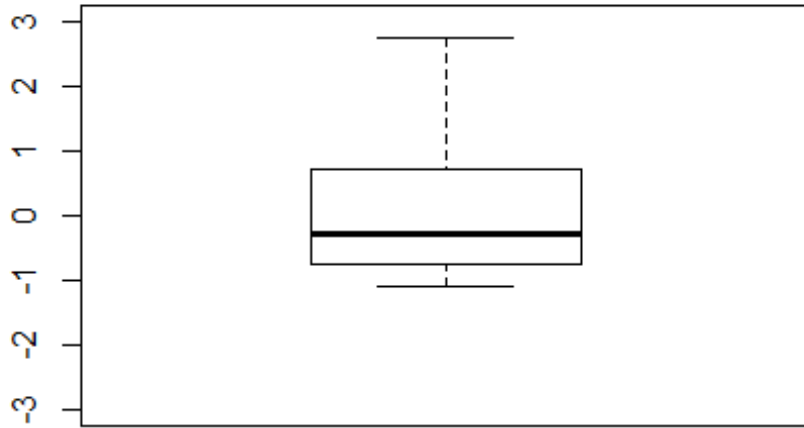
Boxplot du Kurtosis



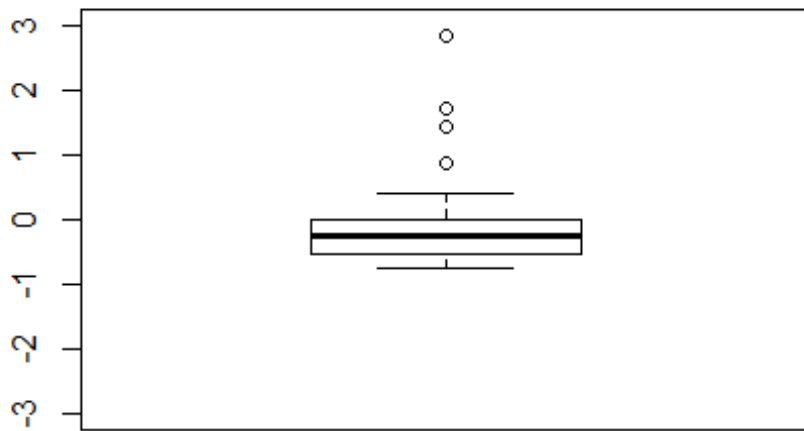
Boxplot de l'intensité moyenne



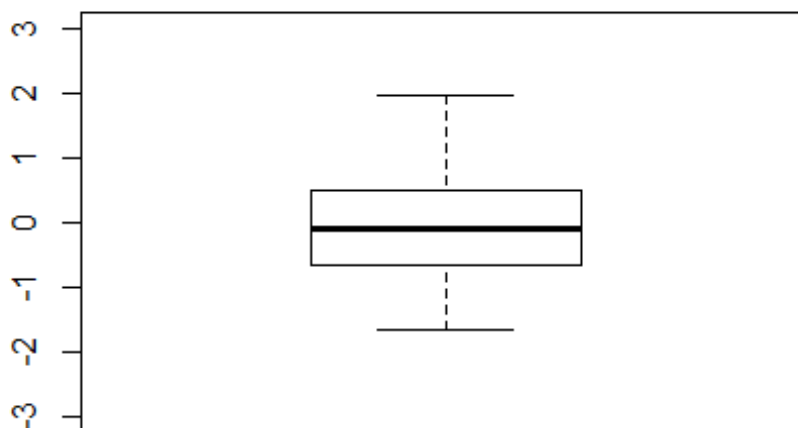
Boxplot de l inverse de la couverture de canopée



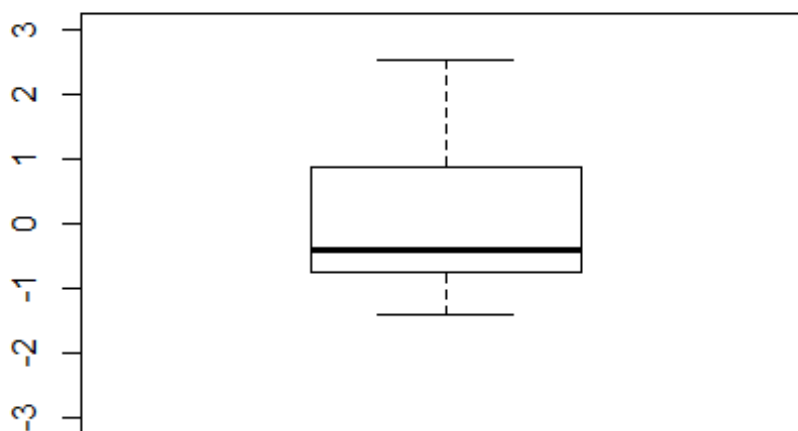
Boxplot du percentile 1%



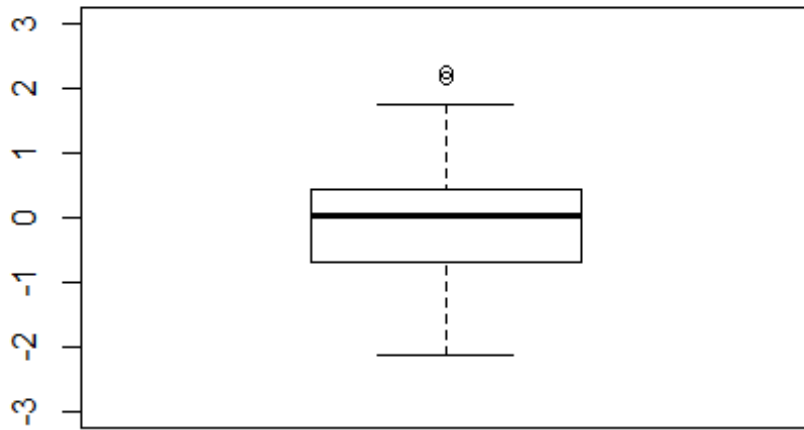
Boxplot du percentile 99%



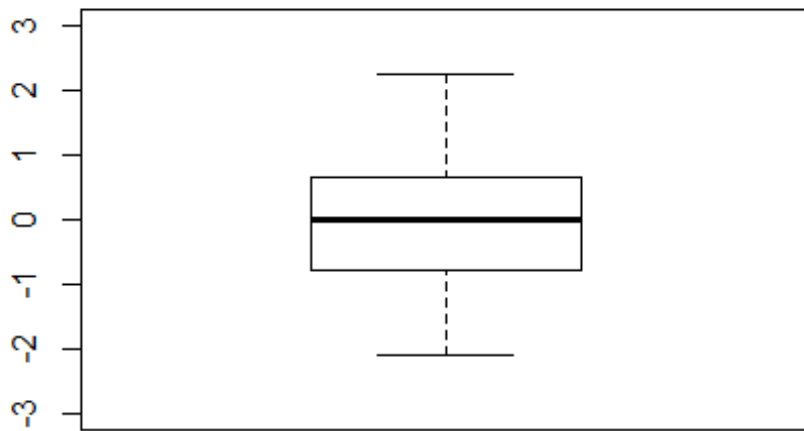
Boxplot du nombre de 5ème retour



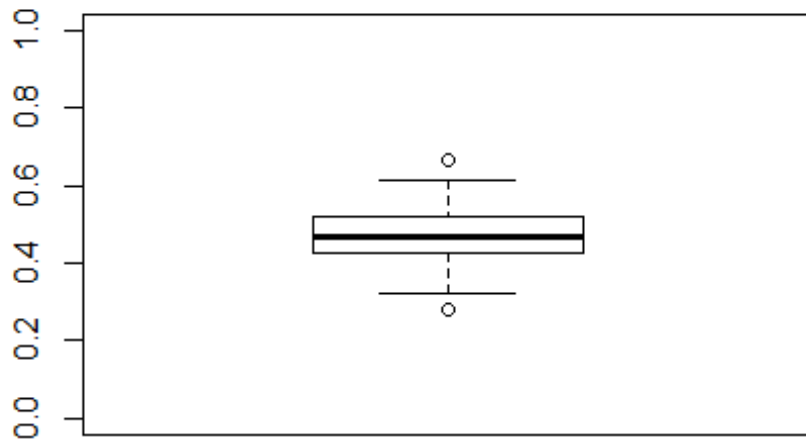
Boxplot de la variance d'intensité



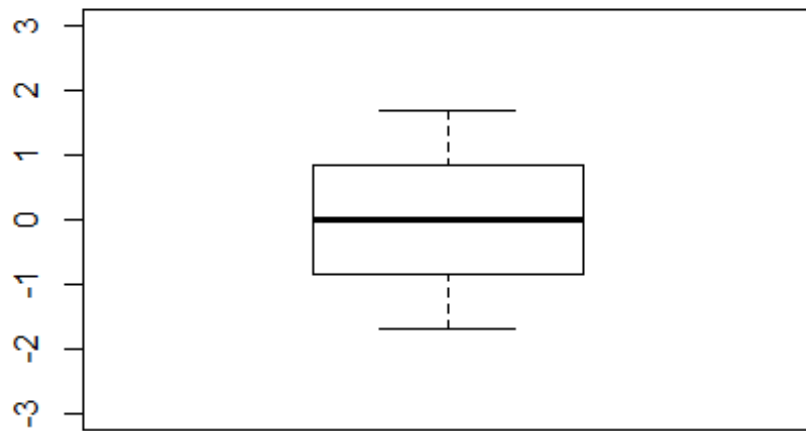
Boxplot du nombre total de retours



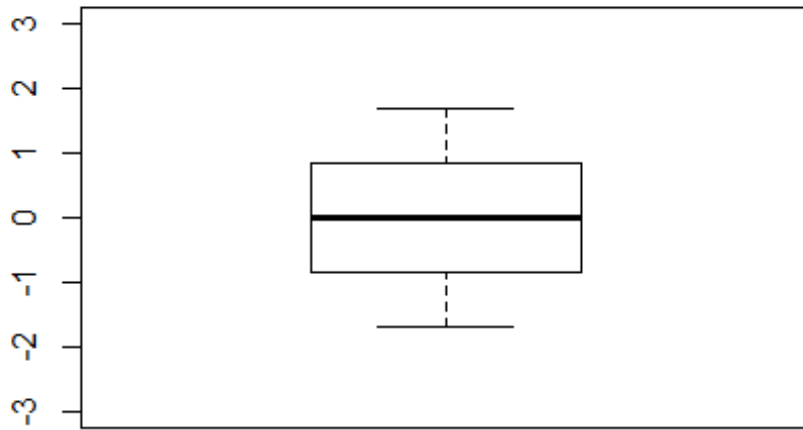
Boxplot du relief de la canopée



Boxplot de la proportion de 1er retours de faible intensité

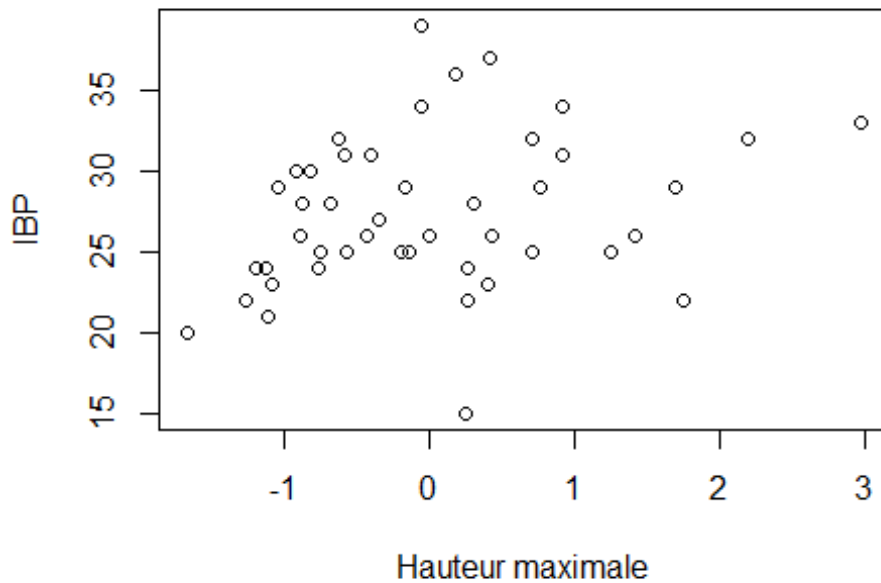


Boxplot du VCI

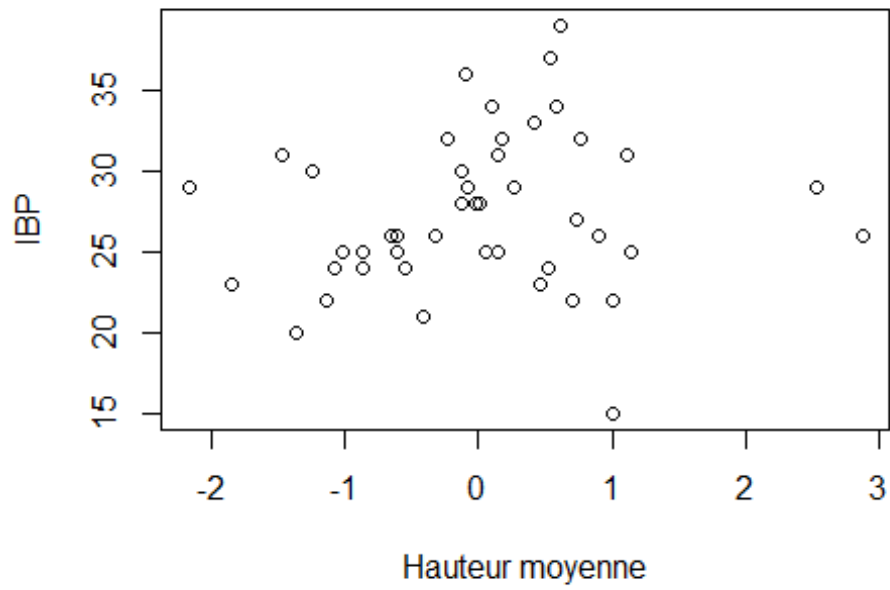


Diagrammes de dispersion de l'IBP en fonction des variables LiDAR non corrélées à l'échelle des plots de 1ha :

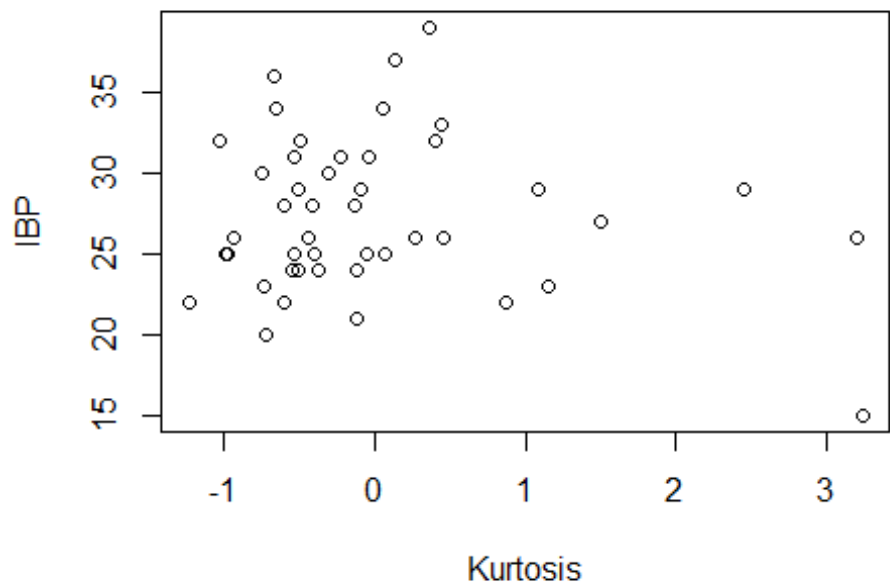
IBP en fonction de la hauteur maximale



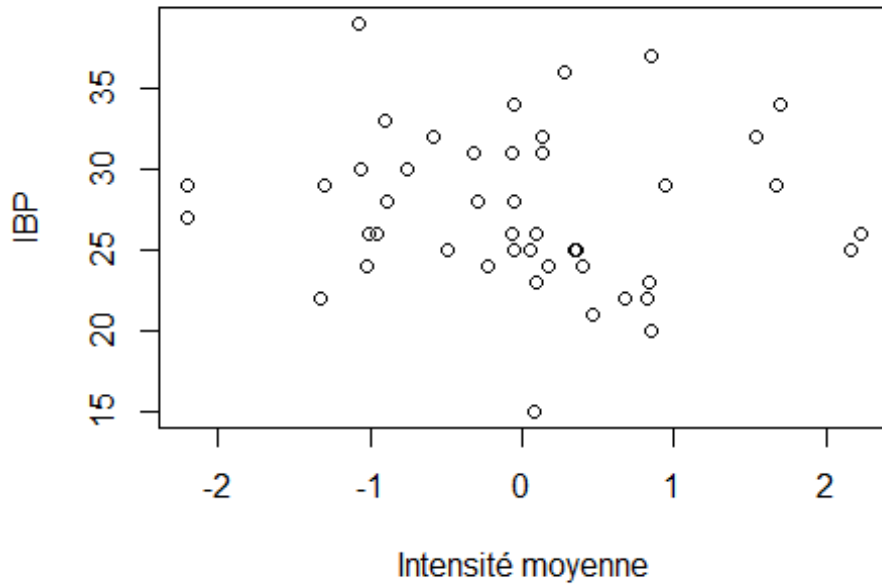
IBP en fonction de la hauteur moyenne



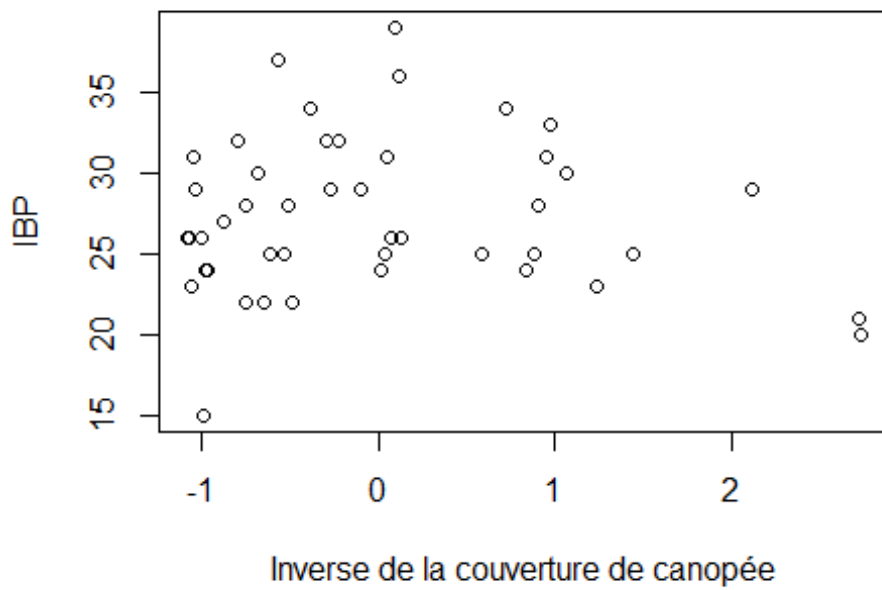
IBP en fonction du Kurtosis



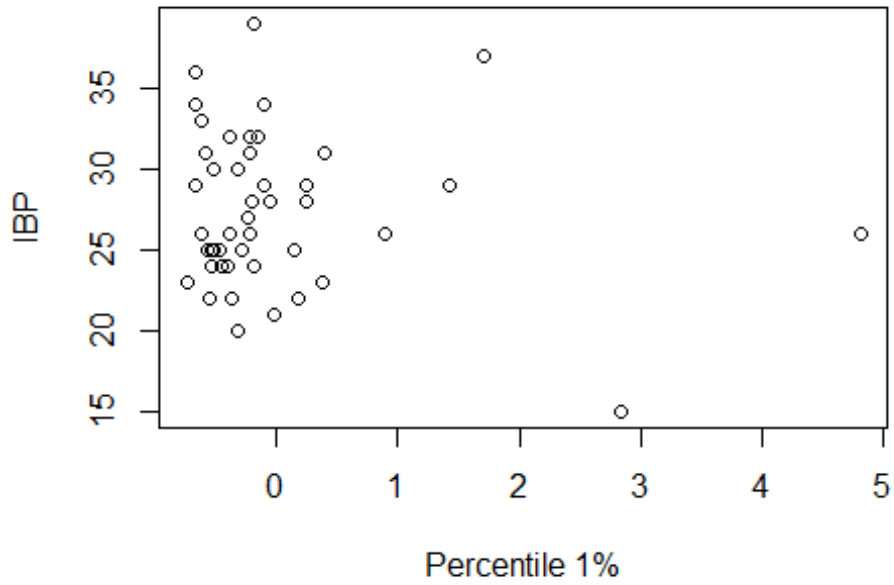
IBP en fonction de l'intensité moyenne



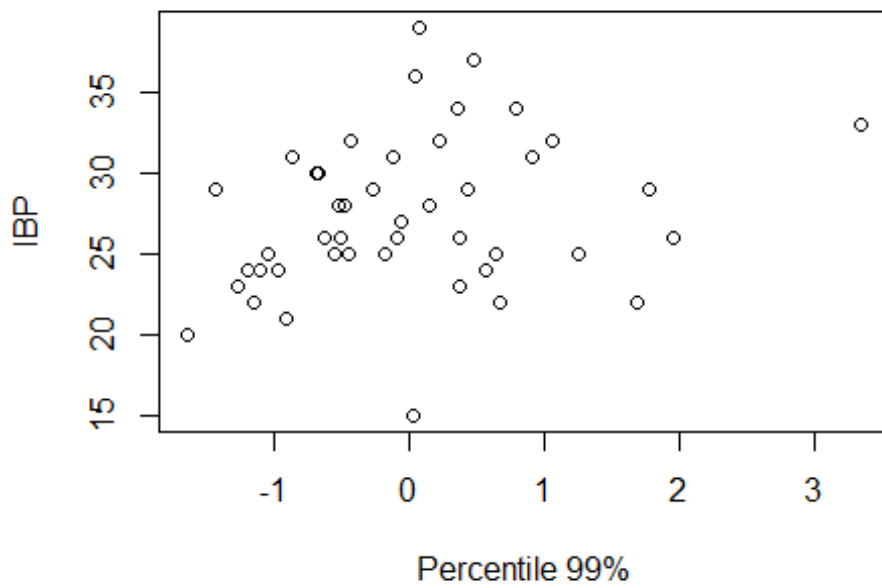
IBP en fonction de l'inverse de la couverture densité de canopée



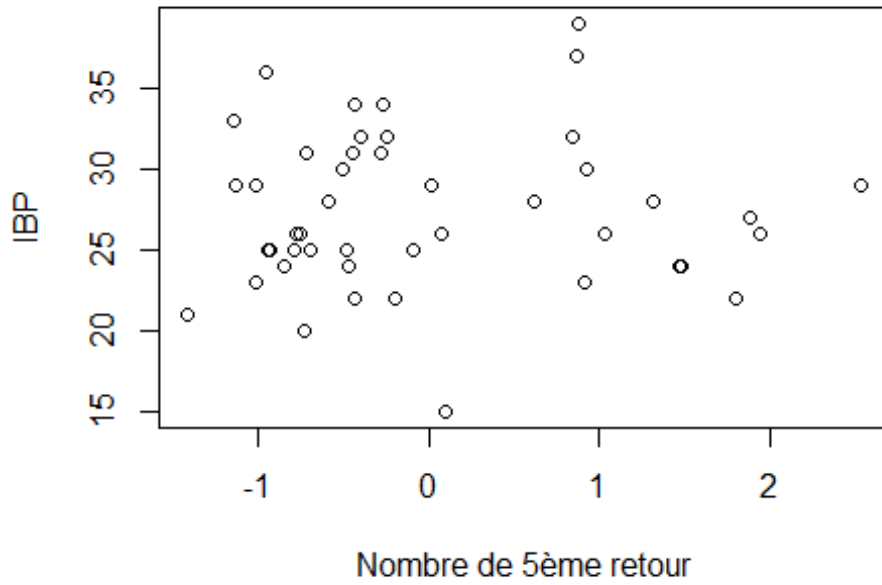
IBP en fonction du percentile 1%



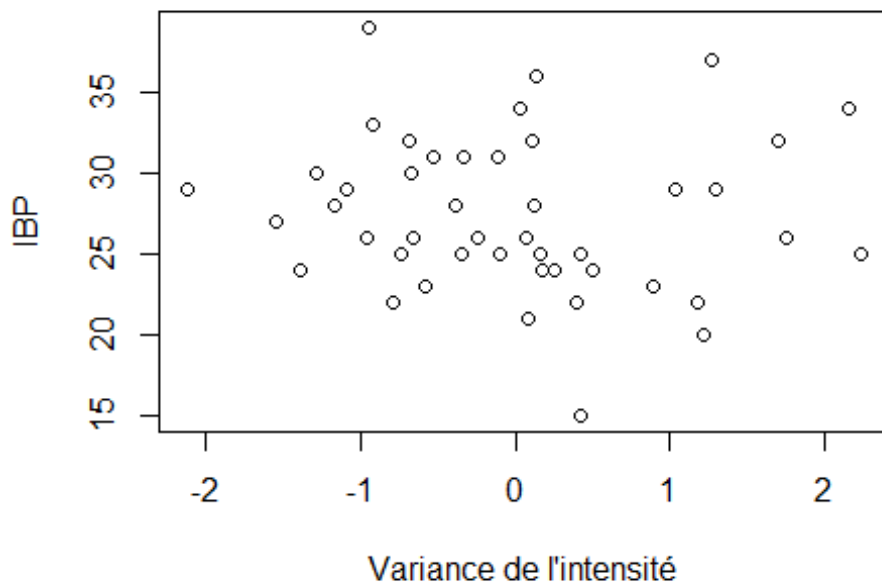
IBP en fonction du percentile 99%



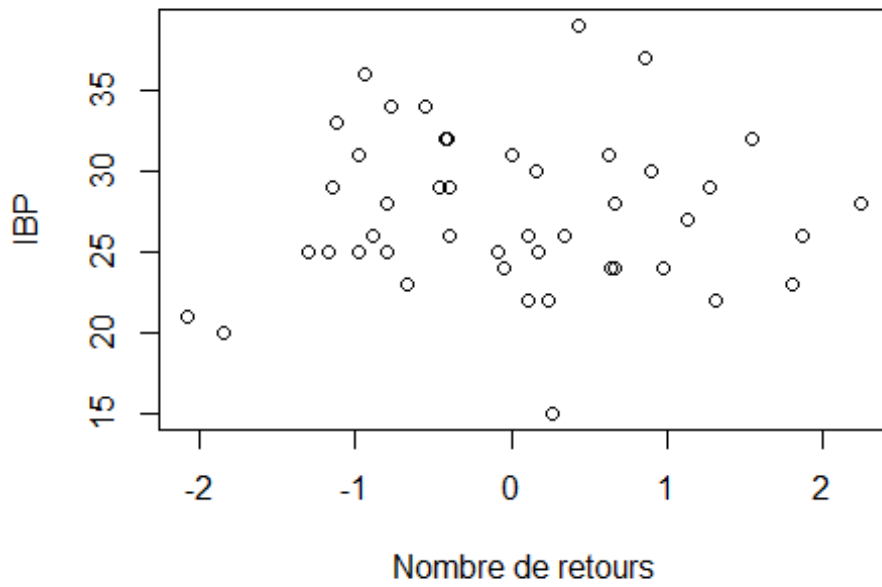
IBP en fonction du nombre de 5ème retour



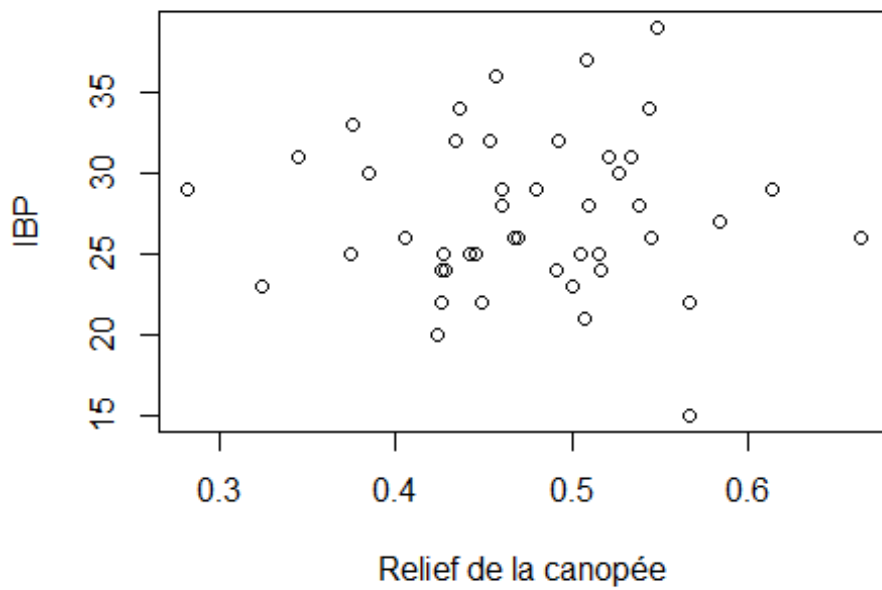
IBP en fonction de la variance de l'intensité



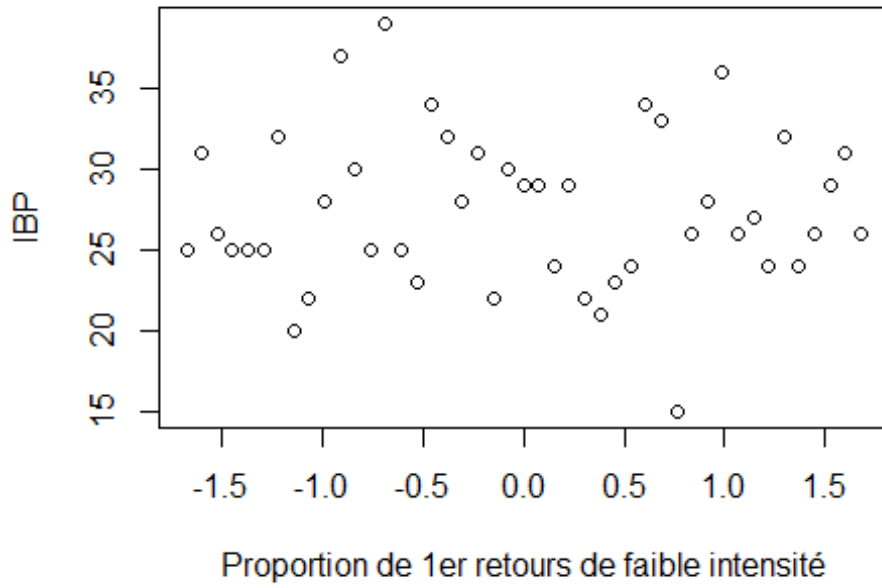
IBP en fonction du nombre de retours



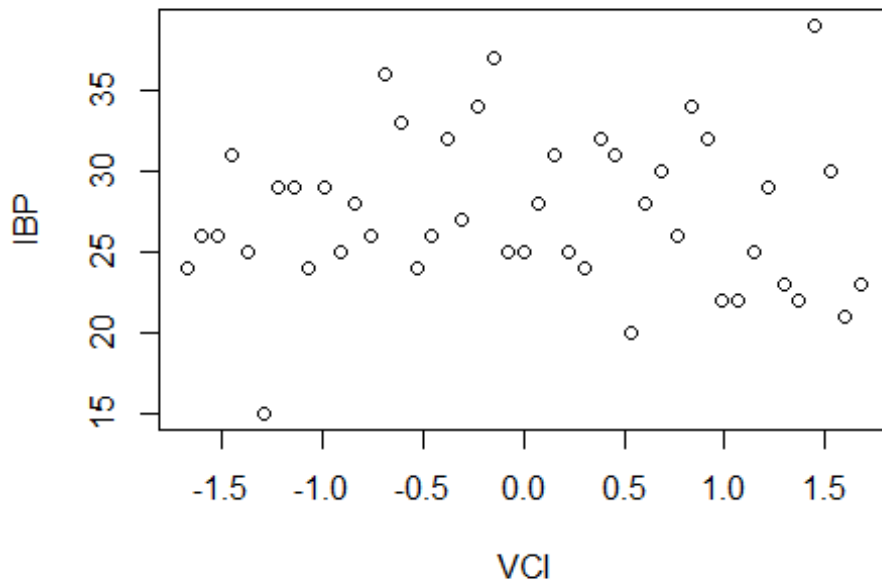
IBP en fonction du relief de la canopée



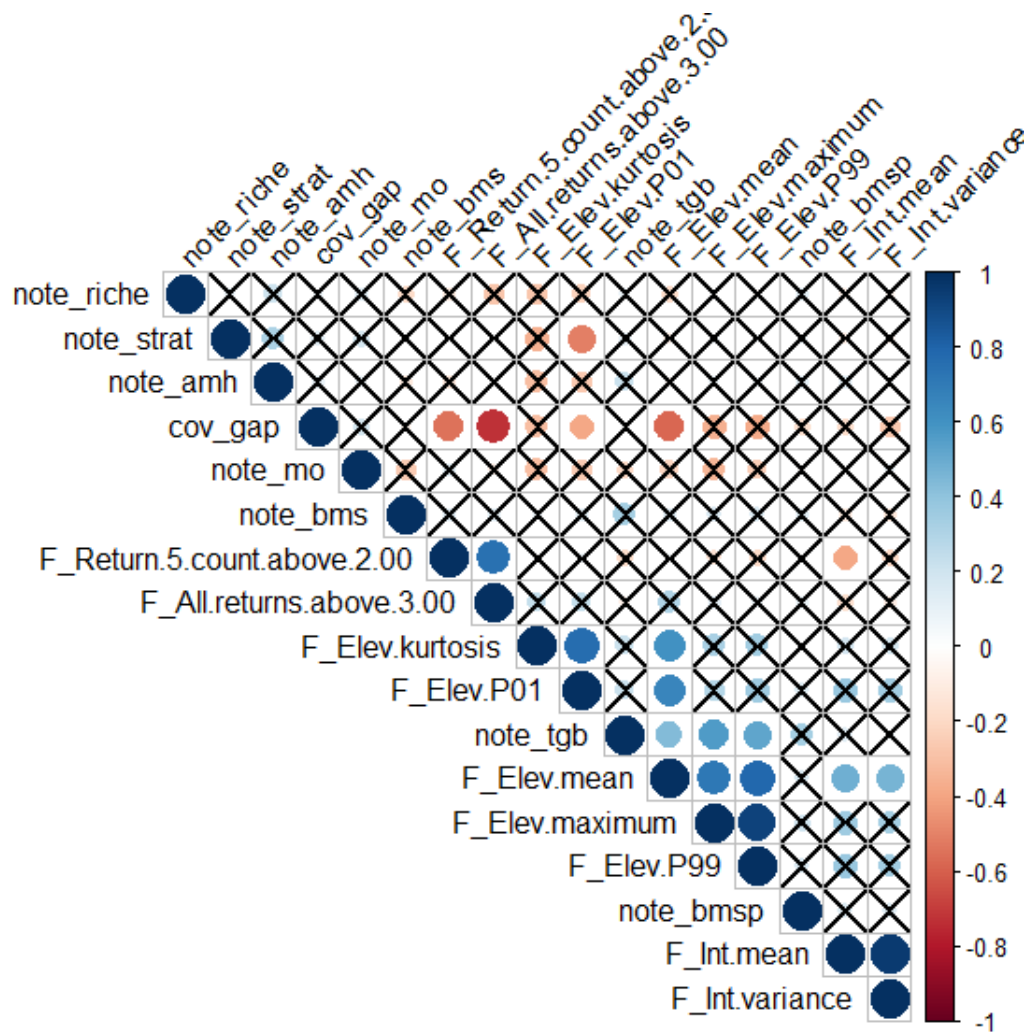
IBP en fonction de la proportion de 1er retours de faible intensité



IBP en fonction du VCI



Corrélogramme des variables LiDAR peu corrélées à l'échelle des plots de 1ha :



Résultats de la modélisation statistique par GLM modèle n°4

```
# Selection du meilleur modèle (AIC le plus faible)
mod_select = glm(note_gest ~ F_Elev.kurtosis+F_Elev.mean,
                 family = "gaussian", data = dataMetricLidIBP)
summary(mod_select)

##
## Call:
## glm(formula = note_gest ~ F_Elev.kurtosis + F_Elev.mean, family =
```

```

"gaussian",
##      data = dataMetricLidIBP)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q        Max
## -7.643  -2.505  -0.199   2.572   9.467
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    20.1510     3.2240   6.250 1.73e-07 ***
## F_Elev.kurtosis -2.7501     1.0895  -2.524  0.0155 *
## F_Elev.mean      0.6517     0.2869   2.271  0.0283 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 16.28713)
##
##      Null deviance: 801.91  on 44  degrees of freedom
## Residual deviance: 684.06  on 42  degrees of freedom
## AIC: 258.17
##
## Number of Fisher Scoring iterations: 2
# Analyse de La deviance du modele
D2 <- (mod_select$null.deviance - mod_select$deviance) /
mod_select$null.deviance
D2

## [1] 0.1469635

```

Résultats de la modélisation statistique par GLM modèle n°5

```

mod_select = glm(formula = note_ibp ~ F_Elev.maximum + F_Elev.kurtosis,
                 family = "gaussian", data = dataMetricLidIBP)
summary(mod_select)

##
## Call:
## glm(formula = note_ibp ~ F_Elev.maximum + F_Elev.kurtosis, family =
## "gaussian",
##      data = dataMetricLidIBP)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q        Max
## -9.1786  -2.8353  -0.3521   2.0914  12.1591
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    17.8763     5.7144   3.128  0.00319 **
## F_Elev.maximum   0.5289     0.2234   2.367  0.02260 *
## F_Elev.kurtosis -1.8108     1.0365  -1.747  0.08795 .
## ---

```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 20.81597)
##
## Null deviance: 1012.80  on 44  degrees of freedom
## Residual deviance:  874.27  on 42  degrees of freedom
## AIC: 269.21
##
## Number of Fisher Scoring iterations: 2

# Analyse de La deviance du modele
D2 <- (mod_select$null.deviance - mod_select$deviance) /
mod_select$null.deviance
D2

## [1] 0.1367785
```

Résultats de la modélisation statistique par GLM modèle n°6

```
mod_select = glm(formula = note_ibp ~
F_Elev.mean+F_Elev.kurtosis+note_ancie+note_habaq,
                 family = "gaussian", data = dataMetricLidIBP)
summary(mod_select)

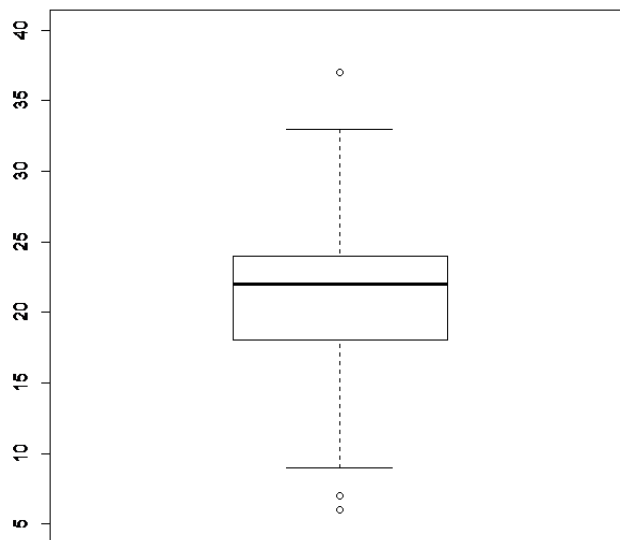
##
## Call:
## glm(formula = note_ibp ~ F_Elev.mean + F_Elev.kurtosis + note_ancie +
## note_habaq, family = "gaussian", data = dataMetricLidIBP)
##
## Deviance Residuals:
##   Min       1Q   Median       3Q      Max
## -7.3678  -1.9506  -0.5232   2.3122   9.5821
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    17.9688     3.6289   4.952 1.38e-05 ***
## F_Elev.mean     0.8104     0.3052   2.655 0.01132 *
## F_Elev.kurtosis -3.1263     1.0958  -2.853 0.00682 **
## note_ancie      1.0640     0.4701   2.263 0.02912 *
## note_habaq      2.0145     0.5690   3.541 0.00103 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 15.83505)
##
## Null deviance: 1012.8  on 44  degrees of freedom
## Residual deviance:  633.4  on 40  degrees of freedom
## AIC: 258.7
##
## Number of Fisher Scoring iterations: 2

# Analyse de La deviance du modele
D2 <- (mod_select$null.deviance - mod_select$deviance) /
```

```
mod_select$null.deviance
D2
## [1] 0.374603
```

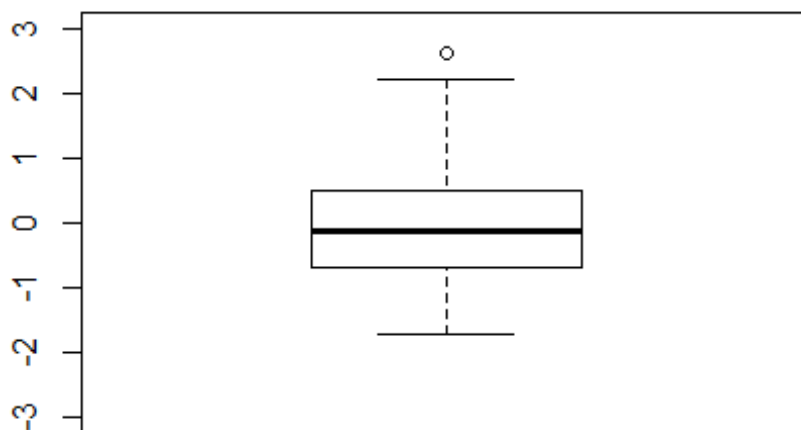
Boîtes à moustaches de la richesse de la strate herbacée à l'échelle des plots de 10m :

Boxplot de la richesse de la strate herbacée

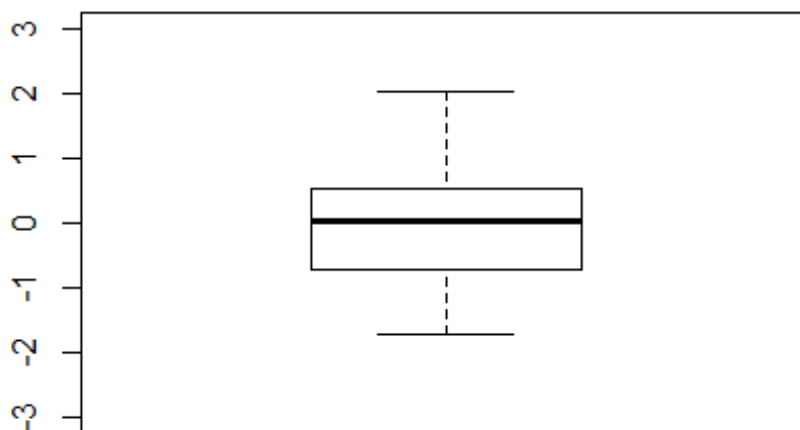


Boîtes à moustaches des variables LiDAR non corrélées à l'échelle des plots de 10m :

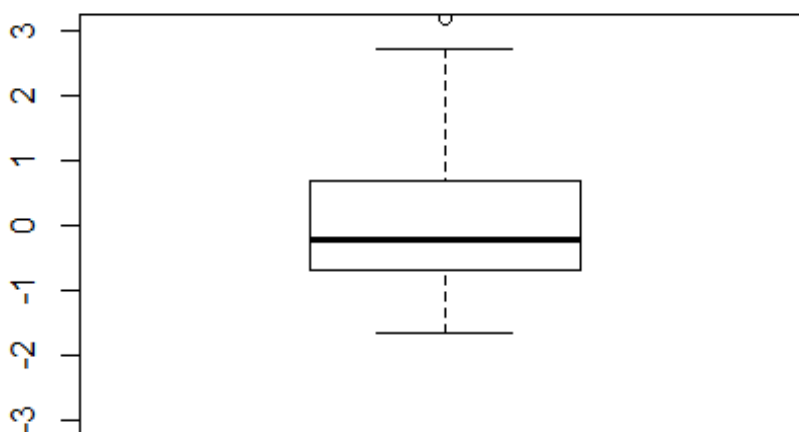
Boxplot du nombre de 5ème retour



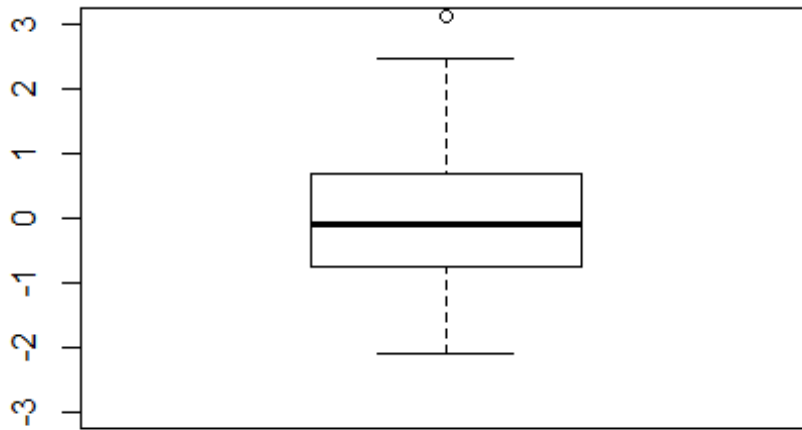
Boxplot de la hauteur maximale



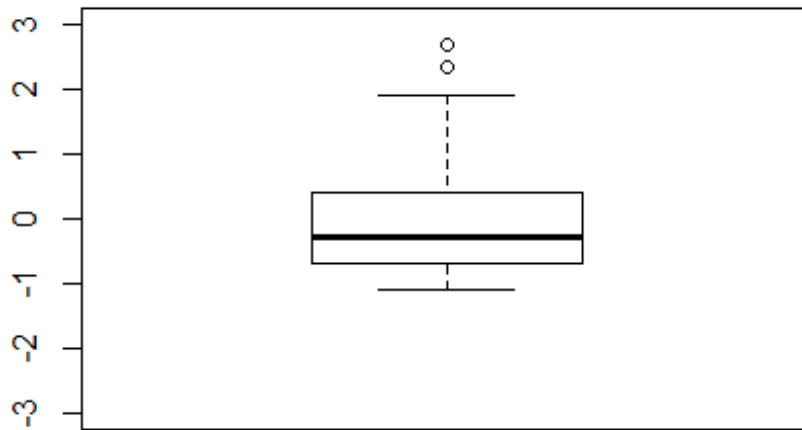
Boxplot de la variance de hauteur



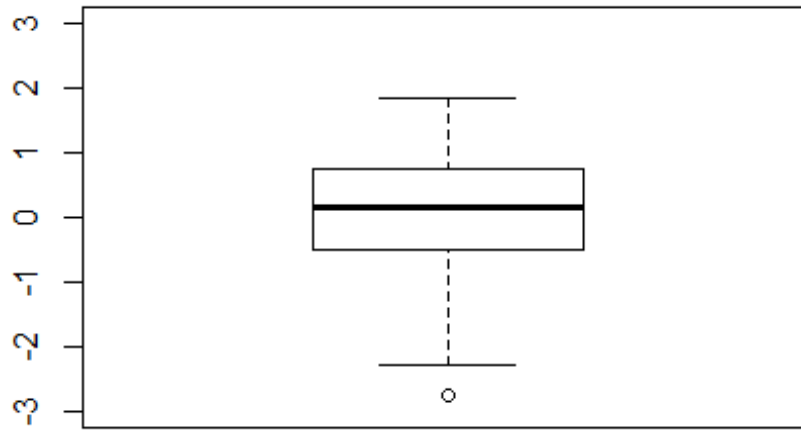
Boxplot du Kurtosis



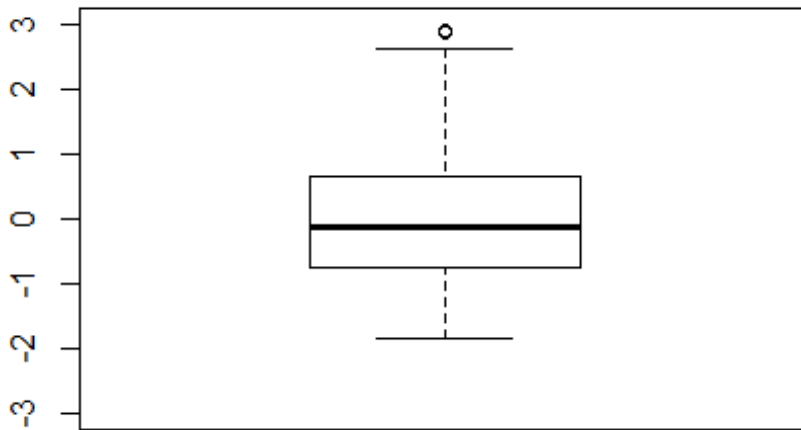
Boxplot du percentile 1%



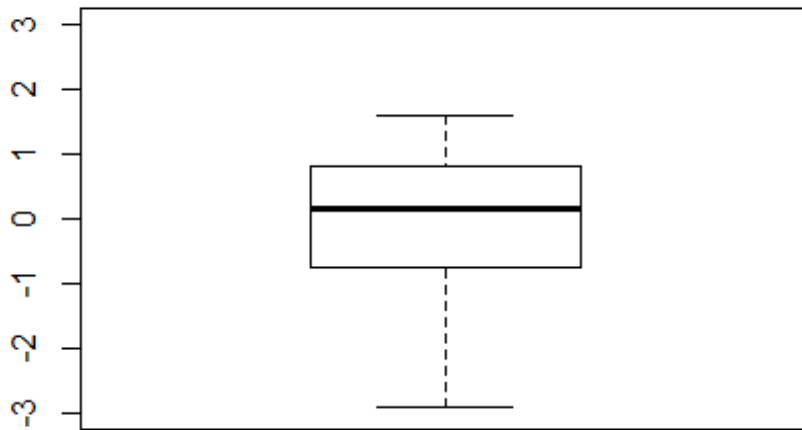
Boxplot du relief de la canopée



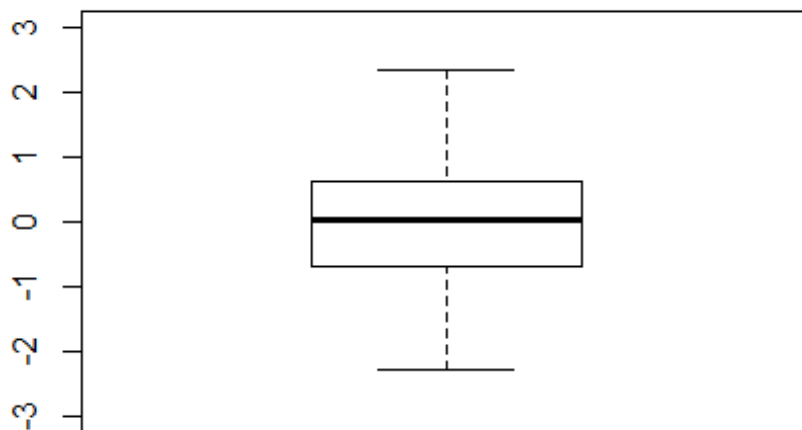
Boxplot de l'intensité maximale



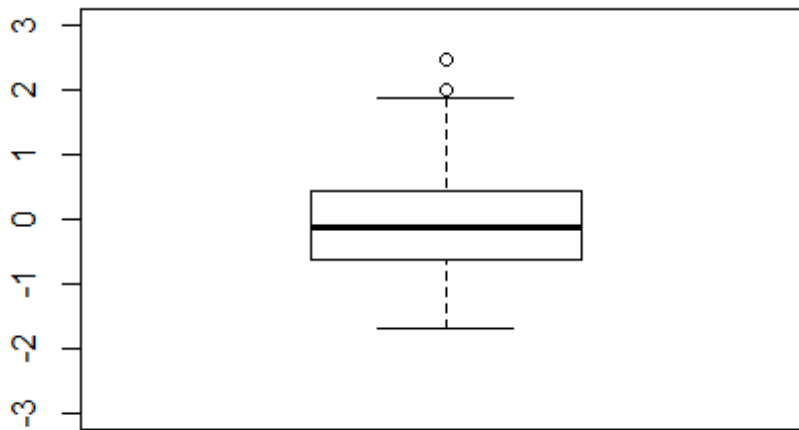
Boxplot des 1ers retours au-dessus de la hauteur moyenne



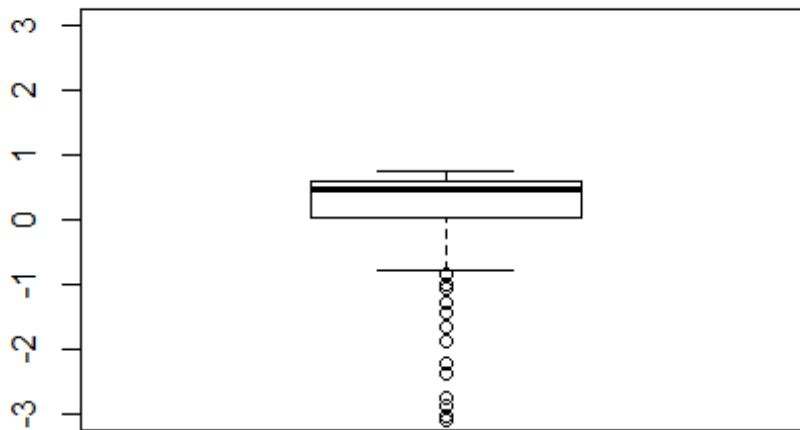
Boxplot des 1ers retours au-dessus du mode de hauteur



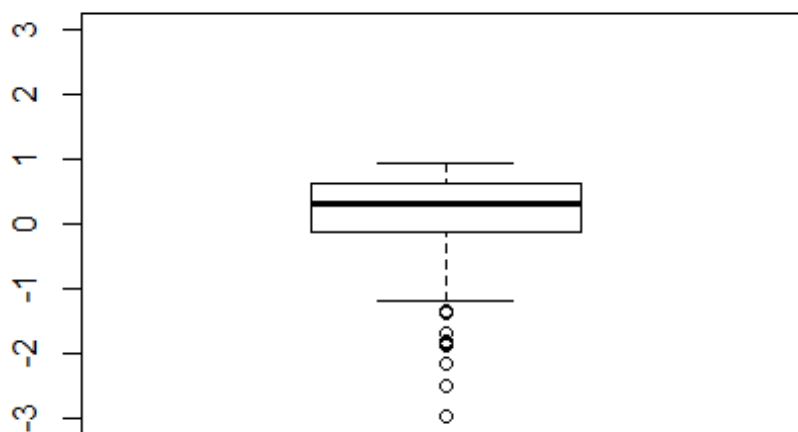
Boxplot de la variance d'intensité



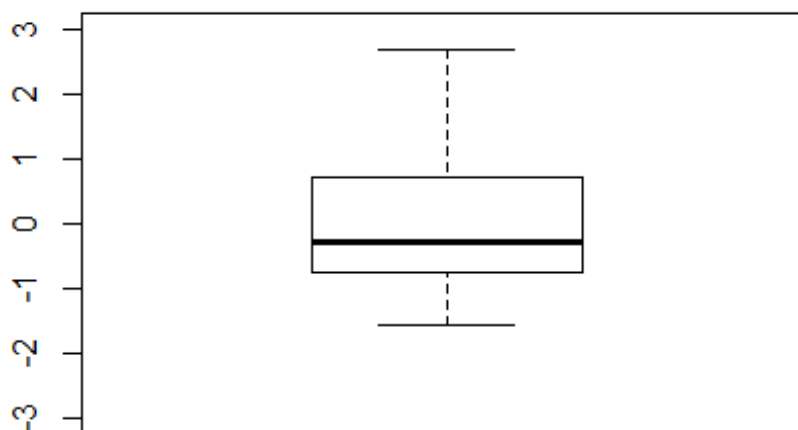
Boxplot de la couverture de canopée à 2m



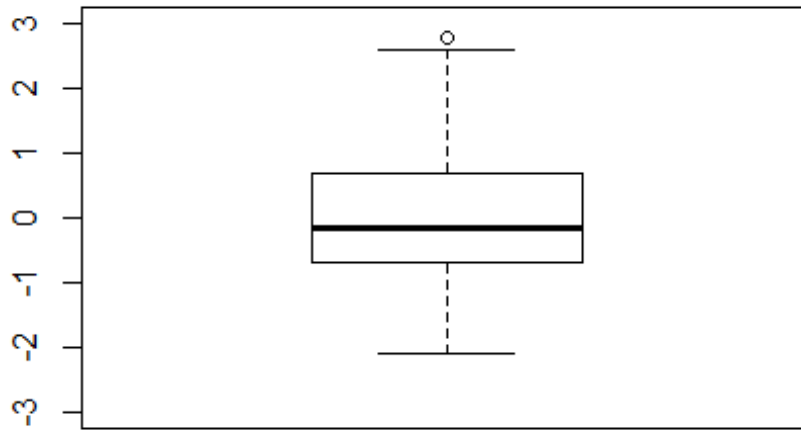
Boxplot de la couverture de canopée à 7m



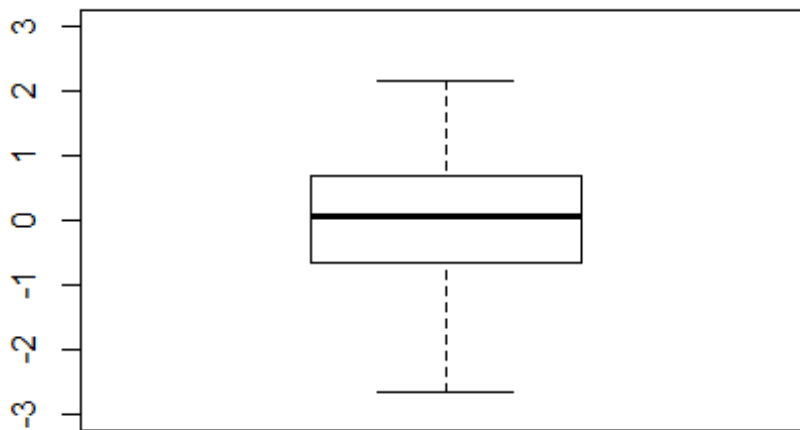
Boxplot de l'inverse de la densité de canopée



Boxplot de la proportion de 1er retours de faible intensité

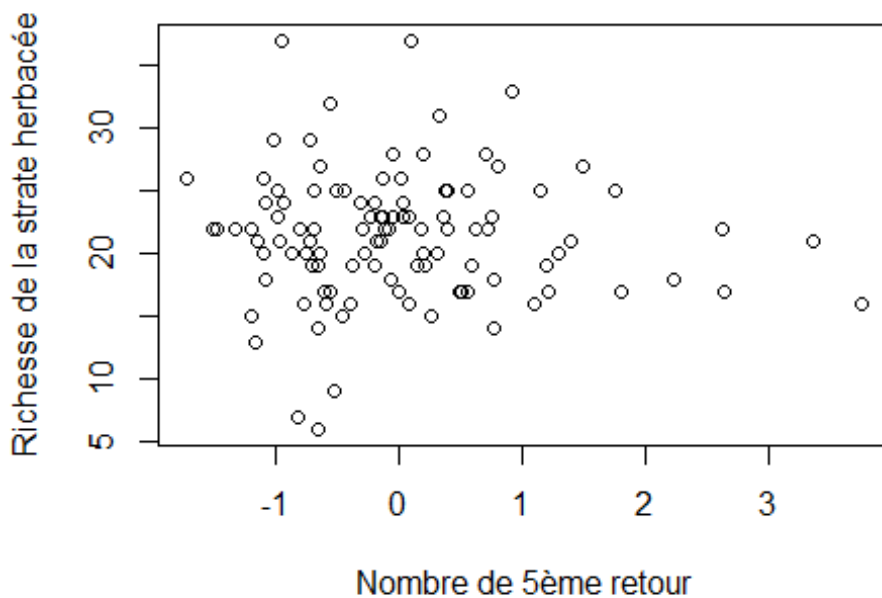


Boxplot du VCI

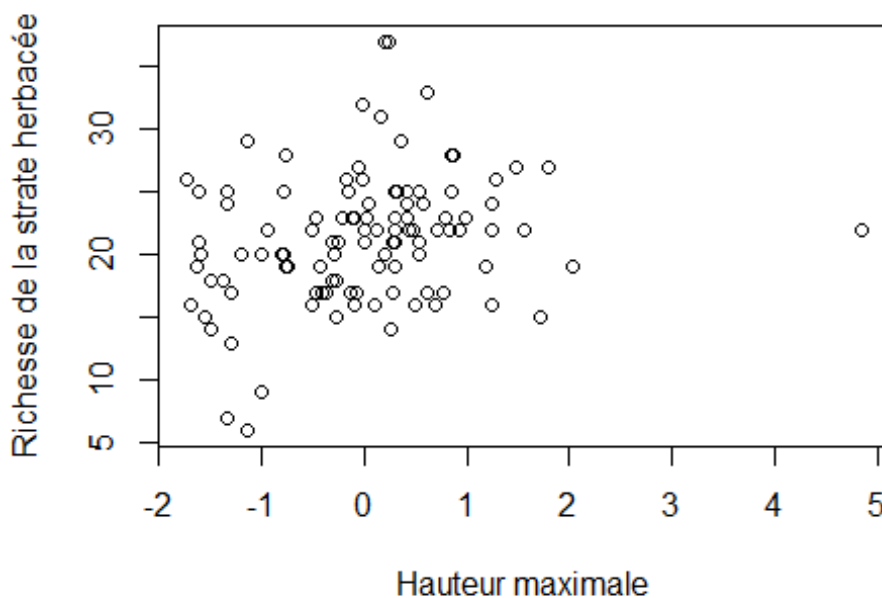


Diagrammes de dispersion de la richesse de la strate herbacée en fonction des variables LiDAR non corrélées à l'échelle des plots de 10m :

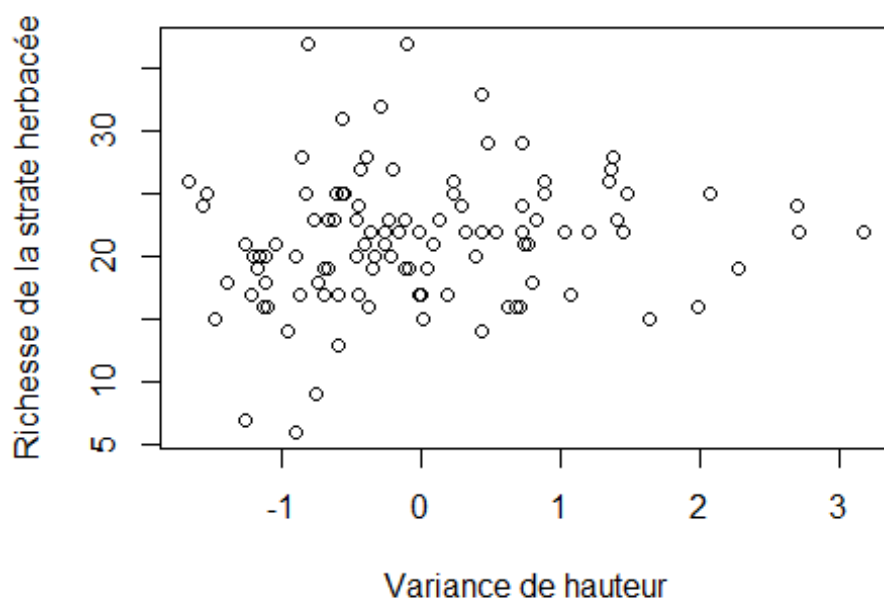
Richesse de la strate herbacée en fonction du nombre de 5ème retour



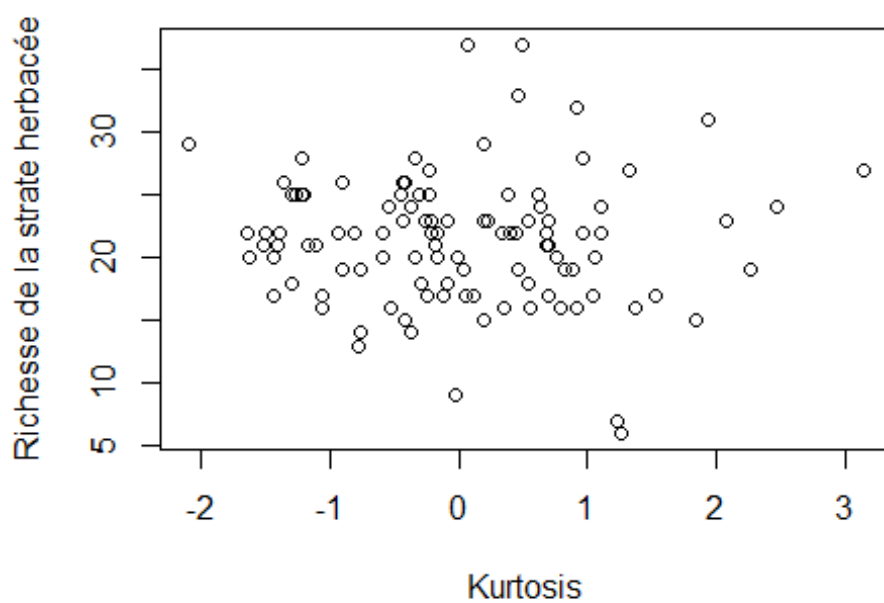
Richesse de la strate herbacée en fonction de la hauteur maximale



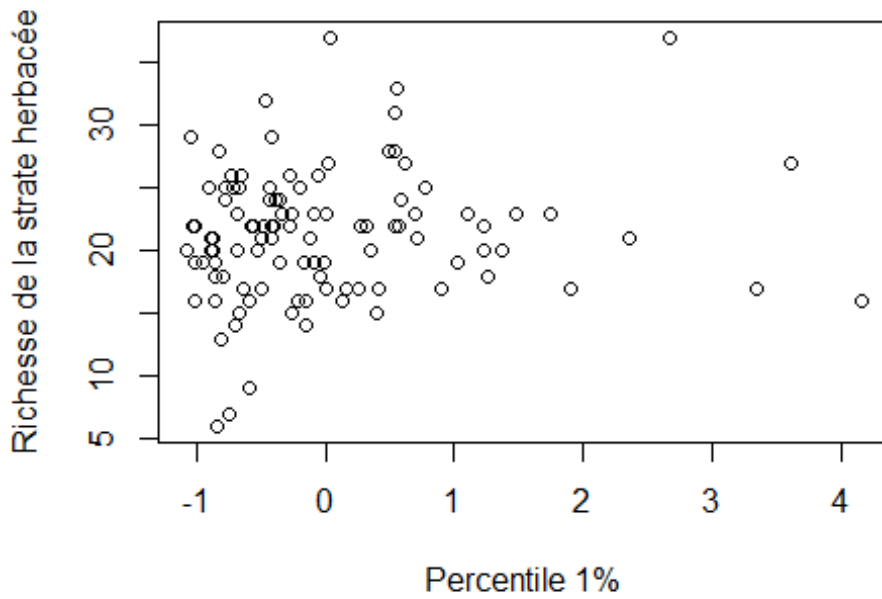
Richesse de la strate herbacée en fonction de la variance de hauteur



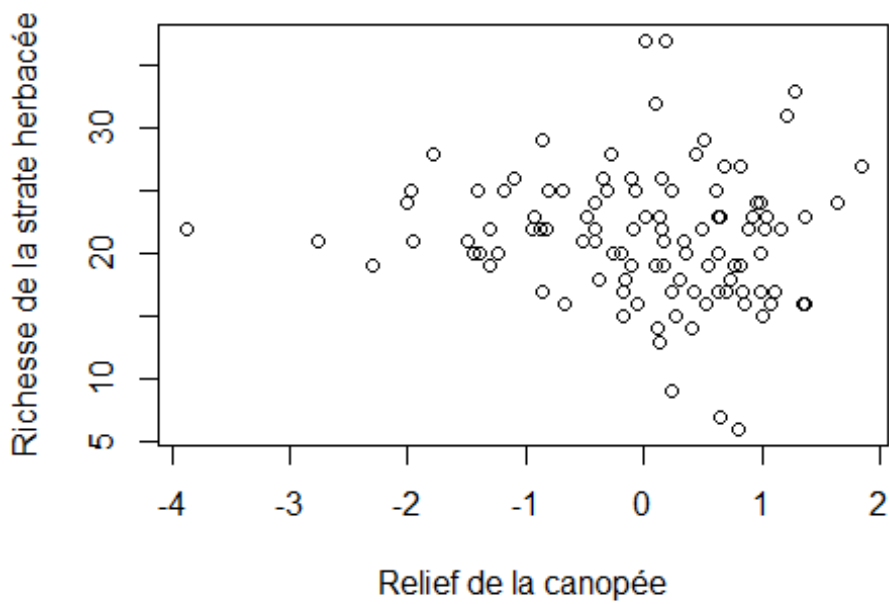
Richesse de la strate herbacée en fonction du Kurtosis



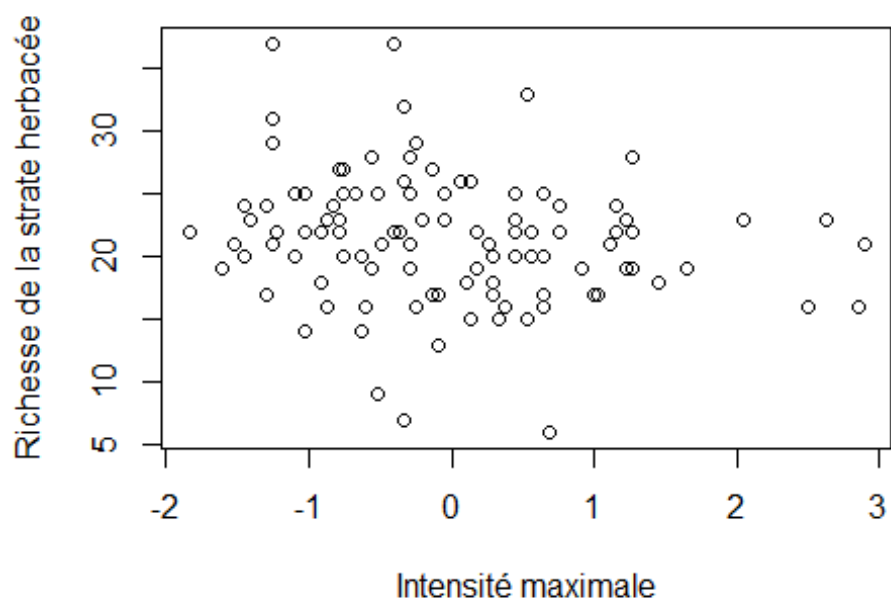
Richesse de la strate herbacée en fonction du percentile 1%



Richesse de la strate herbacée en fonction du relief de la canopée



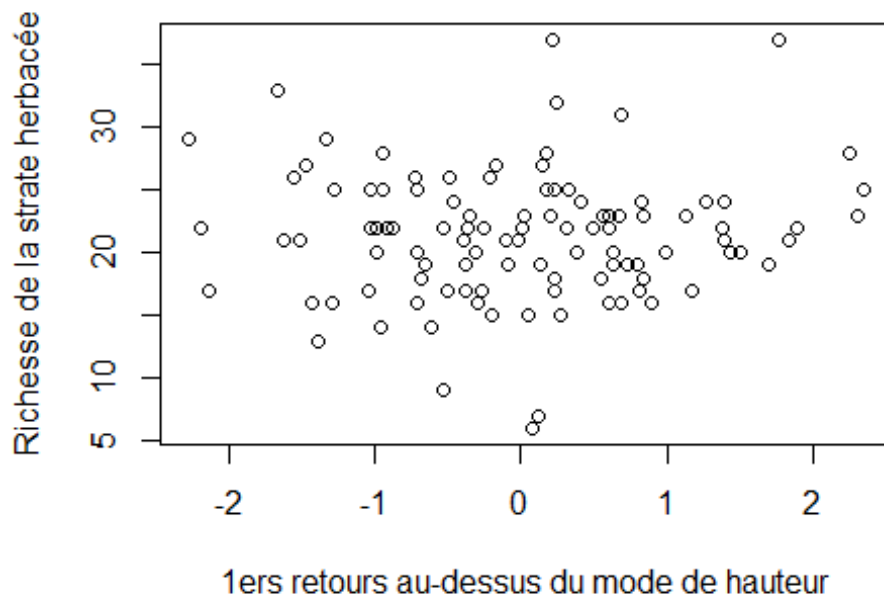
Richesse de la strate herbacée en fonction de l'intensité maximale



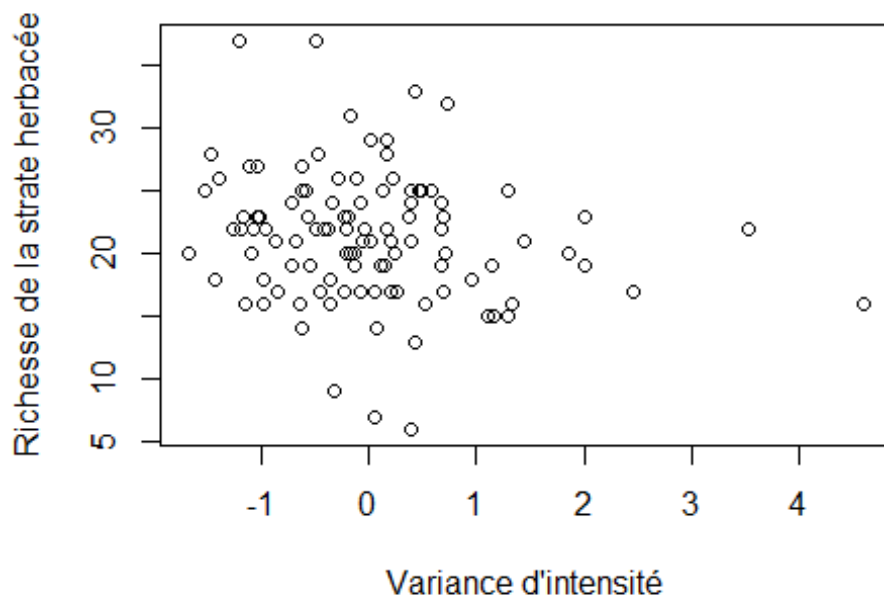
Richesse de la strate herbacée en fonction du nombre de 1ers retours au-dessus de la hauteur moyenne



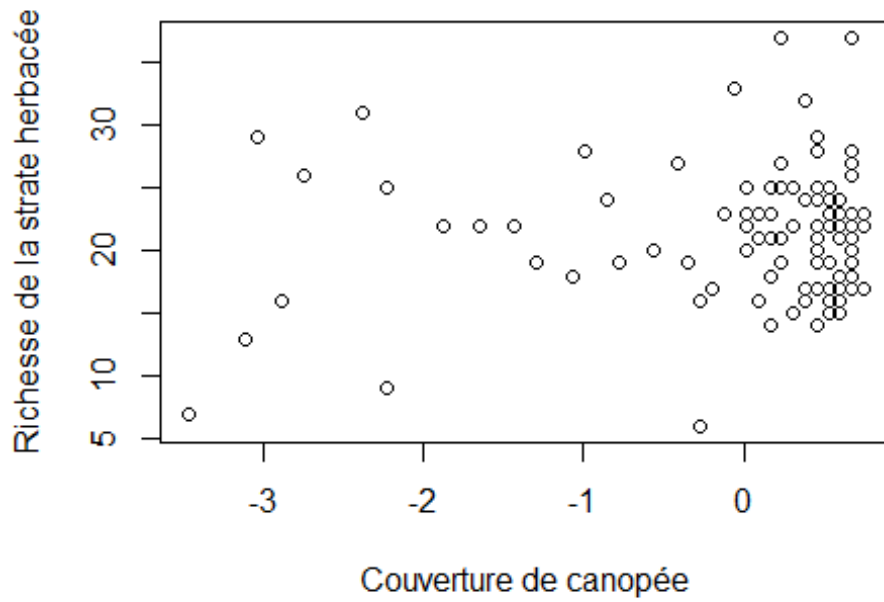
Richesse de la strate herbacée en fonction du nombre de 1ers retours au-dessus du mode de hauteur



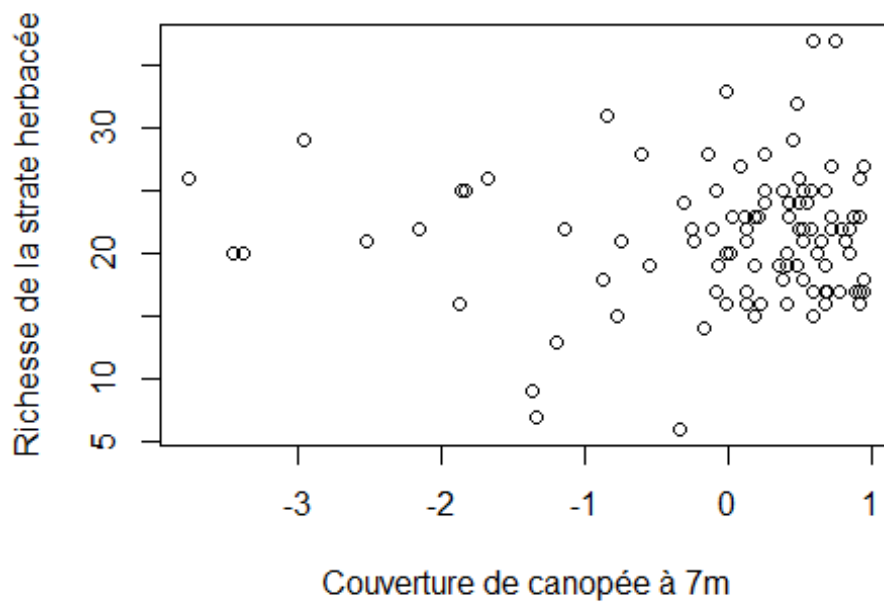
Richesse de la strate herbacée en fonction de la variance d'intensité



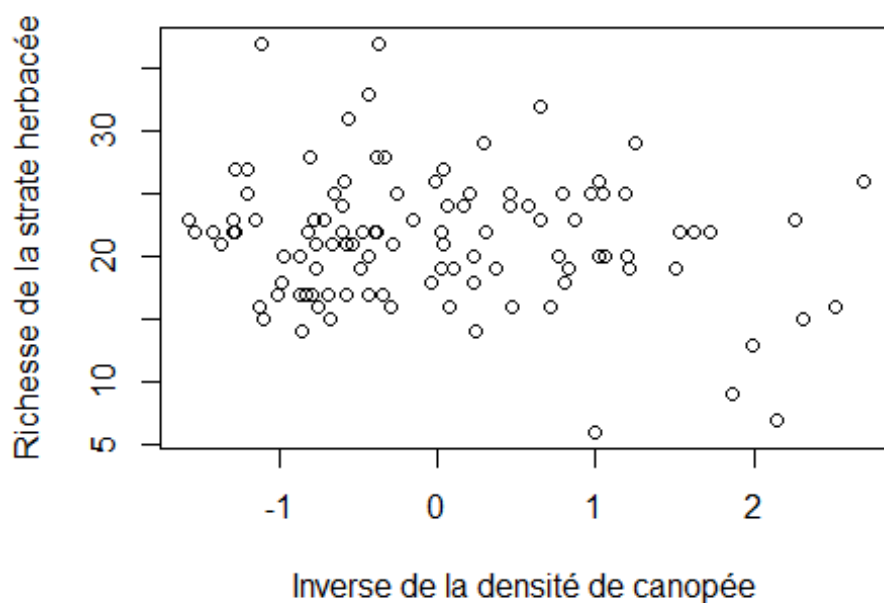
Richesse de la strate herbacée en fonction de la couverture de canopée



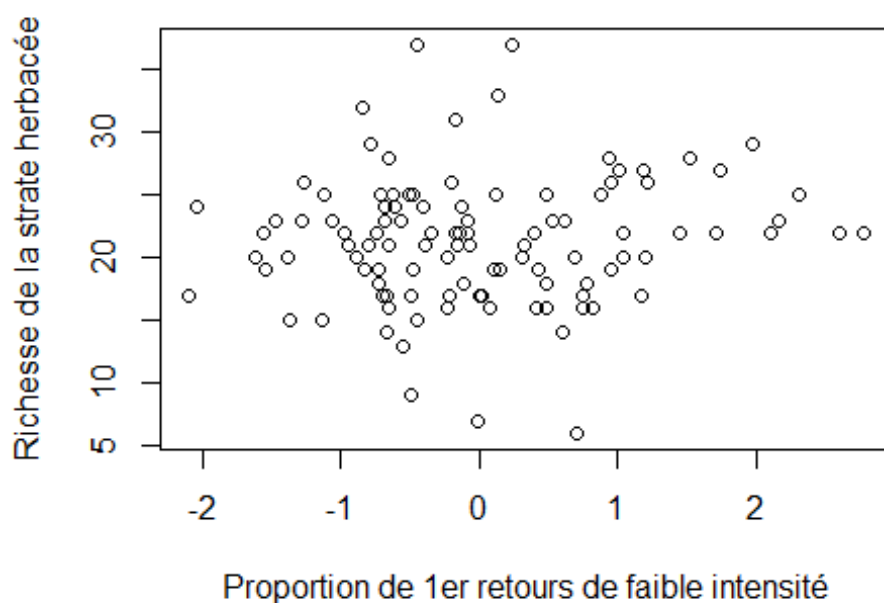
Richesse de la strate herbacée en fonction de e la couverture de canopée à 7m

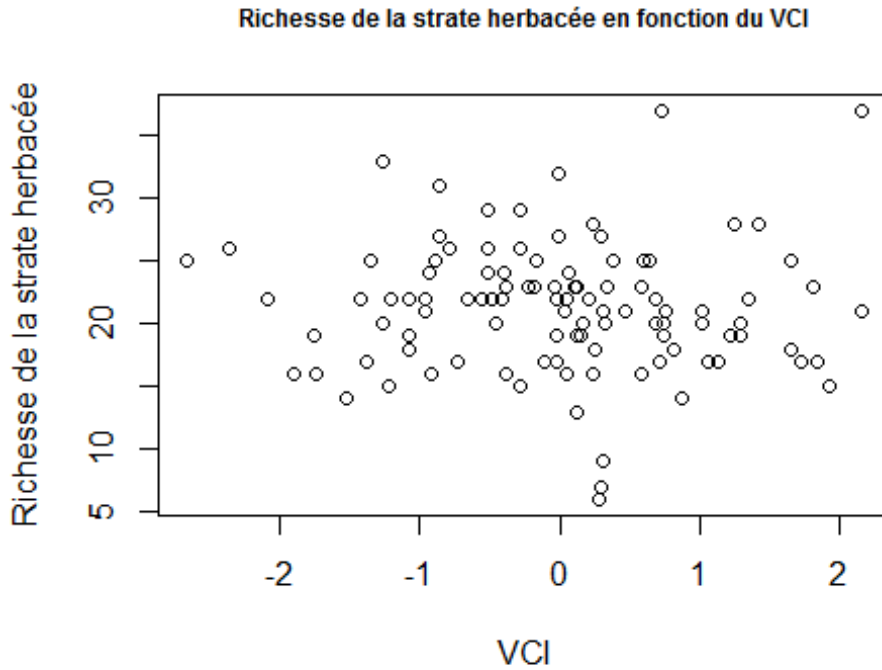


Richesse de la strate herbacée en fonction de l'inverse de la densité de canopée

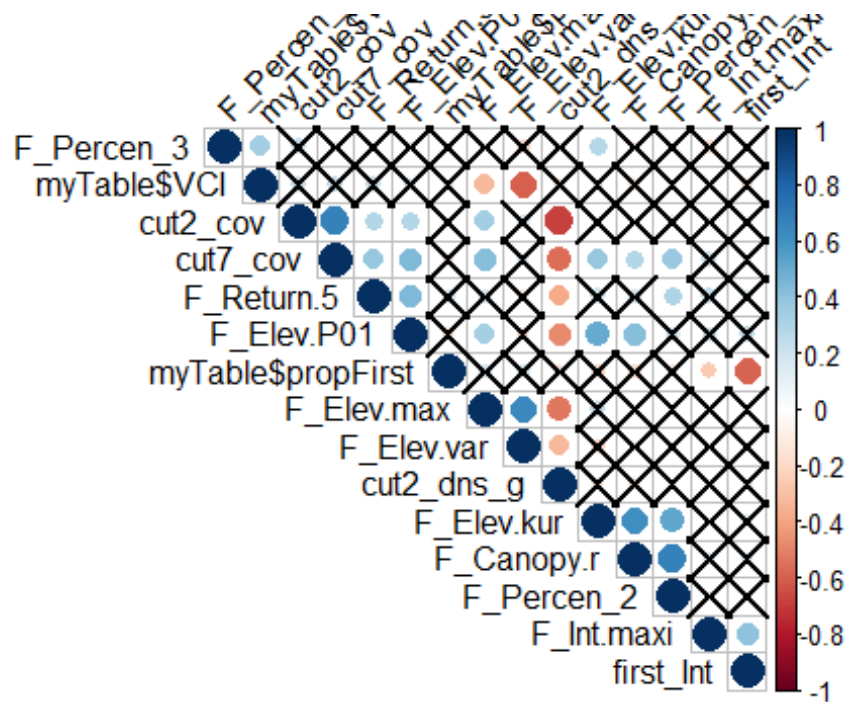


Richesse de la strate herbacée en fonction de la proportion de 1er retours de faible intensité





Corrélogramme des variables LiDAR peu corrélées à l'échelle des plots de 10m :



Résultats de la modélisation statistique par GLM modèle n°16

```
# Selection du meilleur modèle (AIC le plus faible)
```

```
mod_select = glm(dataBota$rich.herb ~ F_Return.5 + F_Elev.max +  
F_Elev.var + F_Elev.kur + F_Elev.P01 + F_Canopy.r + F_Int.maxi + F_Percent_2 +  
F_Percent_3 + first_Int + cut2_cov + cut7_cov +  
+ cut2_dns_g + myTable$propFirst + myTable$VCI, family = "poisson", data =  
dataMetricLidIBPMuesli10mUseful)
```

```
summary(mod_select)
```

Call:

```
glm(formula = dataBota$rich.herb ~ F_Return.5 + F_Elev.max +  
F_Elev.var + F_Elev.kur + F_Elev.P01 + F_Canopy.r + F_Int.maxi +  
F_Percent_2 + F_Percent_3 + first_Int + cut2_cov + cut7_cov +  
cut2_dns_g + myTable$propFirst + myTable$VCI, family = "poisson",  
data = dataMetricLidIBPMuesli10mUseful)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.1839	-0.5999	-0.0477	0.5847	2.9445

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	3.055e+00	2.114e-02	144.517	<2e-16	***
F_Return.5	1.102e-02	3.004e-02	0.367	0.7137	
F_Elev.max	5.111e-02	5.704e-02	0.896	0.3702	
F_Elev.var	2.269e-02	5.496e-02	0.413	0.6797	
F_Elev.kur	-1.518e-02	4.536e-02	-0.335	0.7378	
F_Elev.P01	5.322e-02	4.196e-02	1.269	0.2046	
F_Canopy.r	-4.771e-05	5.593e-02	-0.001	0.9993	
F_Int.maxi	-3.964e-02	2.540e-02	-1.560	0.1186	
F_Percent_2	-4.162e-02	3.529e-02	-1.180	0.2382	
F_Percent_3	-3.470e-03	2.616e-02	-0.133	0.8945	
first_Int	-6.344e-02	3.343e-02	-1.897	0.0578	.
cut2_cov	-1.527e-03	3.779e-02	-0.040	0.9678	
cut7_cov	-1.912e-02	3.788e-02	-0.505	0.6137	
cut2_dns_g	5.423e-03	3.932e-02	0.138	0.8903	
myTable\$propFirst	-4.159e-02	3.253e-02	-1.279	0.2010	
myTable\$VCI	6.755e-03	3.328e-02	0.203	0.8391	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 137.02 on 105 degrees of freedom
Residual deviance: 113.93 on 90 degrees of freedom
AIC: 662.47

Number of Fisher Scoring iterations: 4

```

> # Analyse de la deviance du modele
> D2 <- (mod_select$null.deviance - mod_select$deviance) /
  mod_select$null.deviance
> D2
[1] 0.168491

```

Résultats de la modélisation statistique par GLMM modèle n°16

```

# Selection du meilleur modèle
# Percentage first returns above mean = F_Percen_2
mod_select = glmer(formula = dataBota$rich.herb ~ F_Percen_2 + (1 |
as.factor(dataBota$id_bois)),
  family = "poisson", data = dataMetricLidIBPMuesli10m,
na.action = "na.fail")
summary(mod_select)

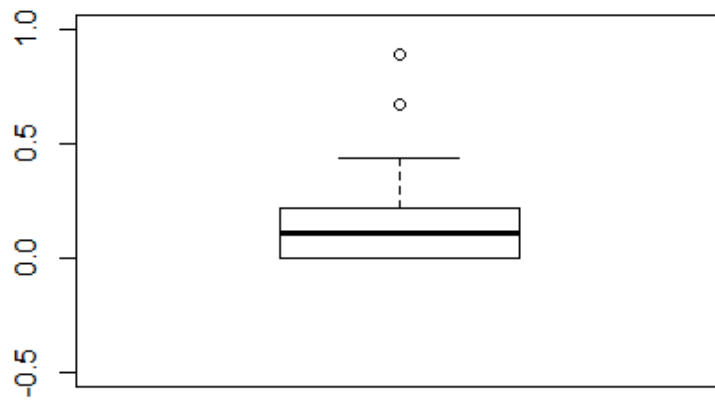
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: poisson ( log )
## Formula:
## dataBota$rich.herb ~ F_Percen_2 + (1 | as.factor(dataBota$id_bois))
## Data: dataMetricLidIBPMuesli10m
##
##      AIC      BIC   logLik deviance df.resid
##  632.2   640.2  -313.1   626.2     103
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.97455 -0.59379 -0.00528  0.49994  2.32483
##
## Random effects:
## Groups                Name      Variance Std.Dev.
## as.factor(dataBota$id_bois) (Intercept) 0.02441  0.1562
## Number of obs: 106, groups: as.factor(dataBota$id_bois), 29
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.552776   0.276958  12.83  <2e-16 ***
## F_Percen_2  -0.006763   0.003675  -1.84  0.0657 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
r.squaredGLMM(mod_select)

## The result is correct only if all data used by the model has not changed
## since model was fitted.
##
##      R2m      R2c
## 0.02863417 0.36301710

```

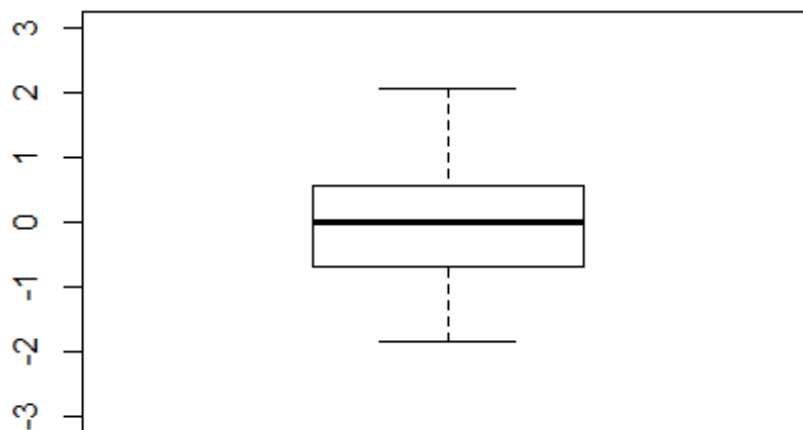
Boîtes à moustaches de la prédation des chenilles à l'échelle des plots de 26m :

Boxplot de la prédation des défoliateurs

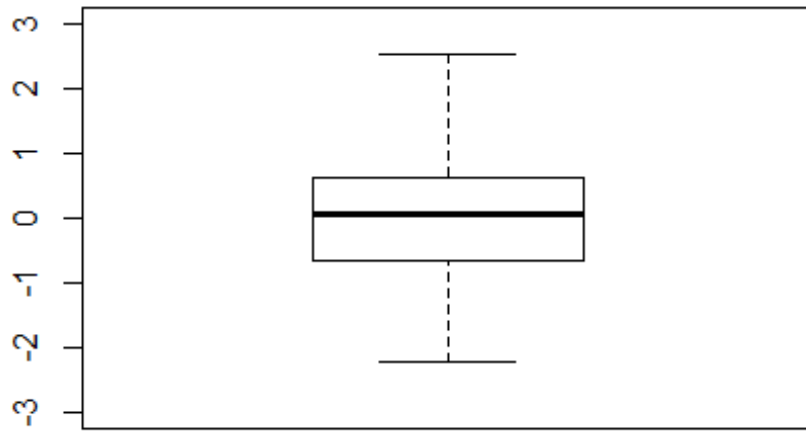


Boîtes à moustaches des variables LIDAR non corrélées à l'échelle des plots de 26m :

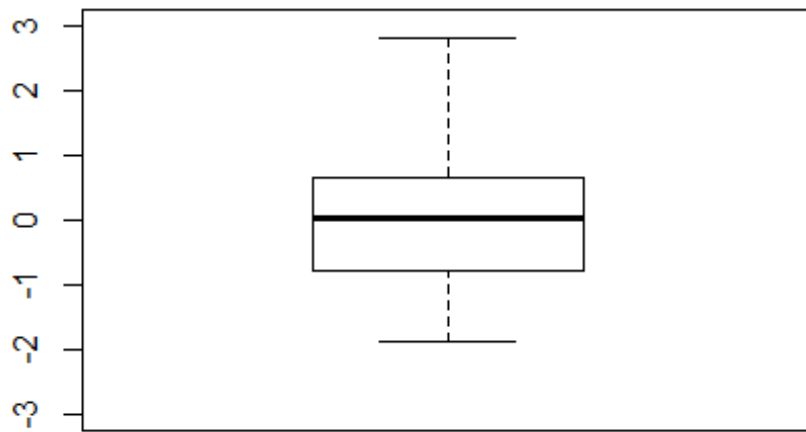
Boxplot de la hauteur maximale



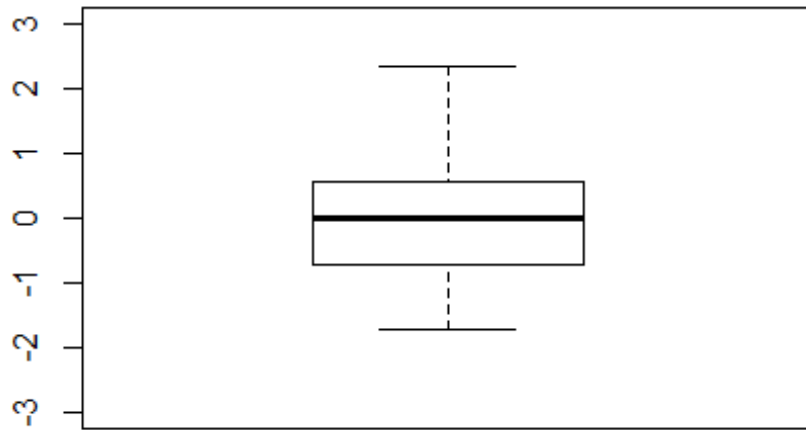
Boxplot de la hauteur moyenne



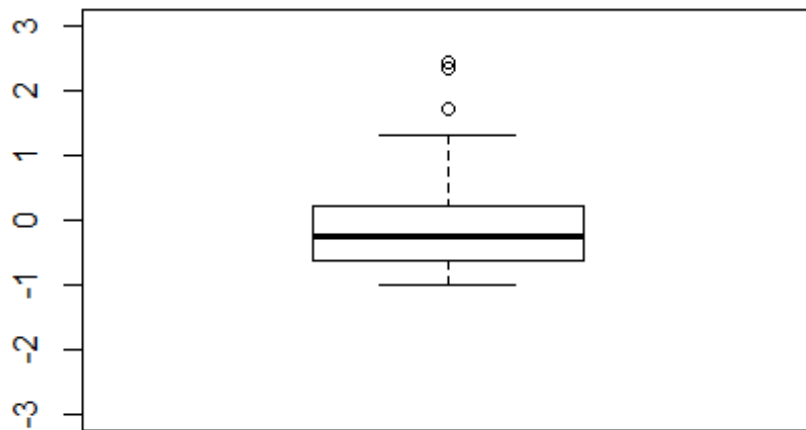
Boxplot du Kurtosis



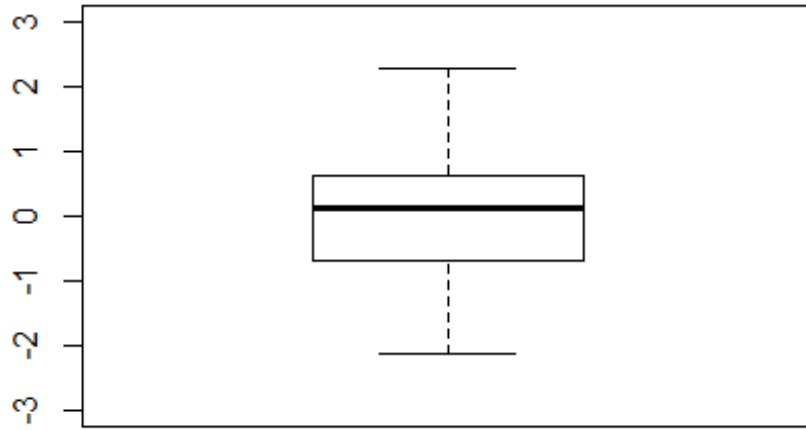
Boxplot du Skewness



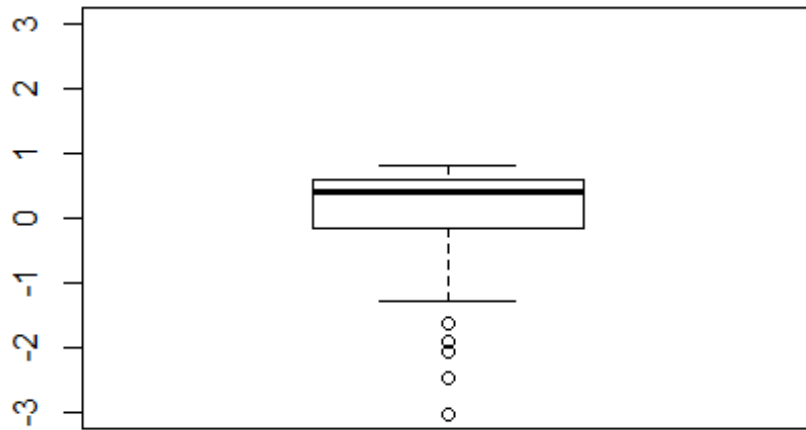
Boxplot du percentile 1%



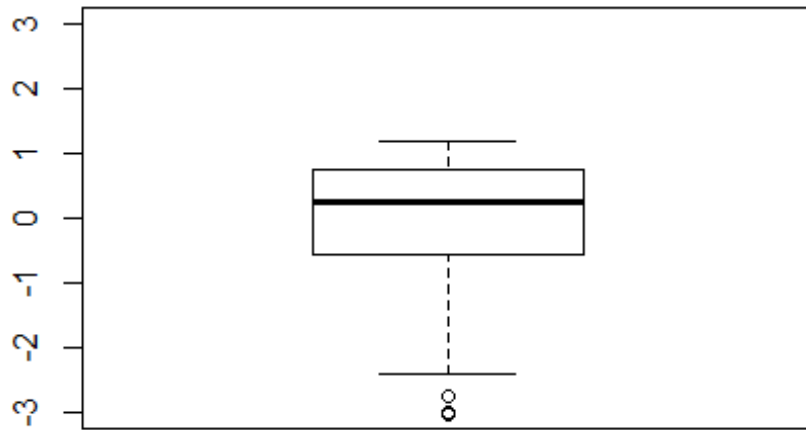
Boxplot du percentile 99%



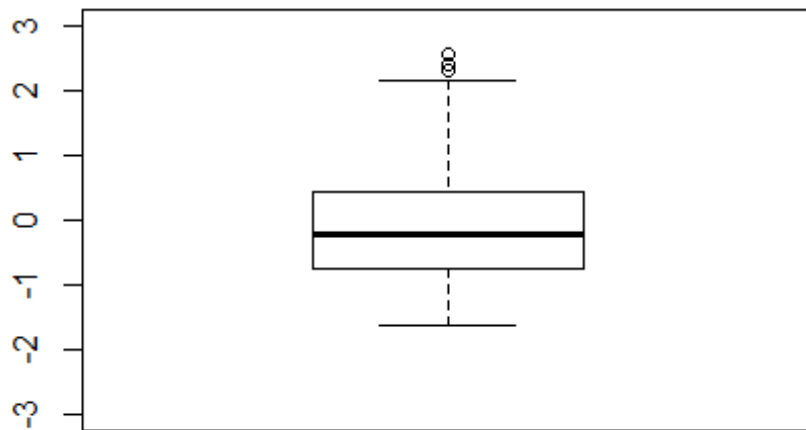
Boxplot de la couverture de canopée



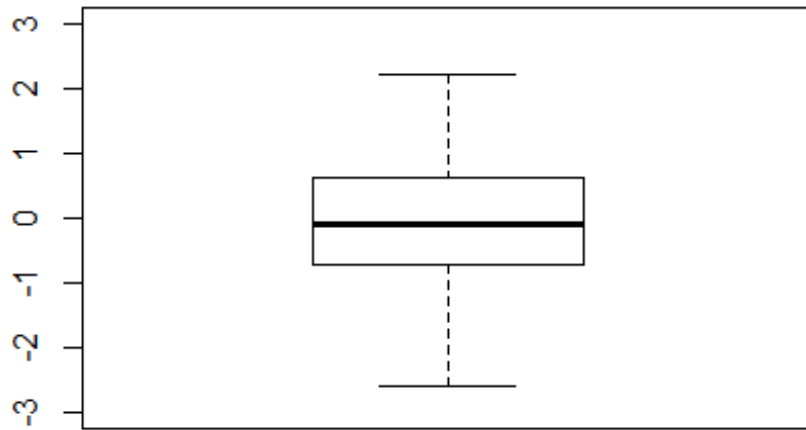
Boxplot de la couverture de canopée



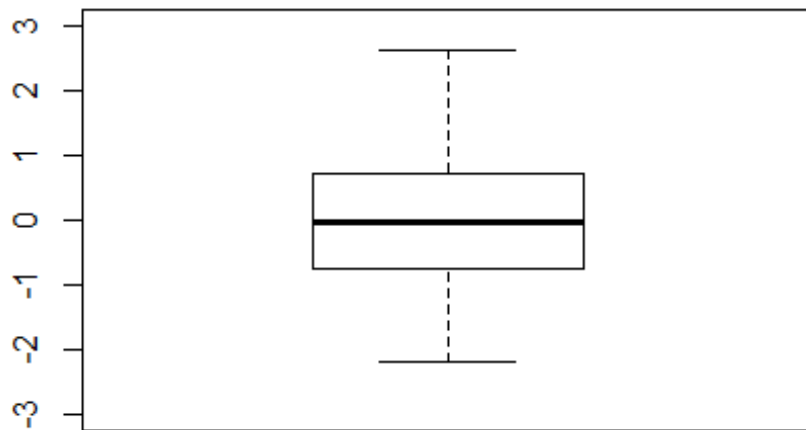
Boxplot du nombre total de retours



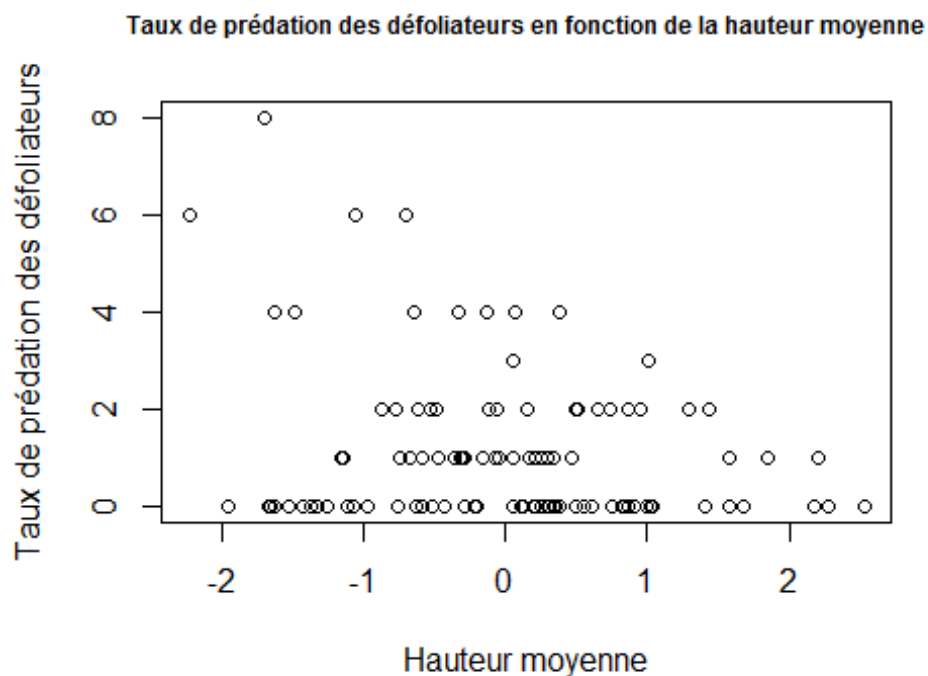
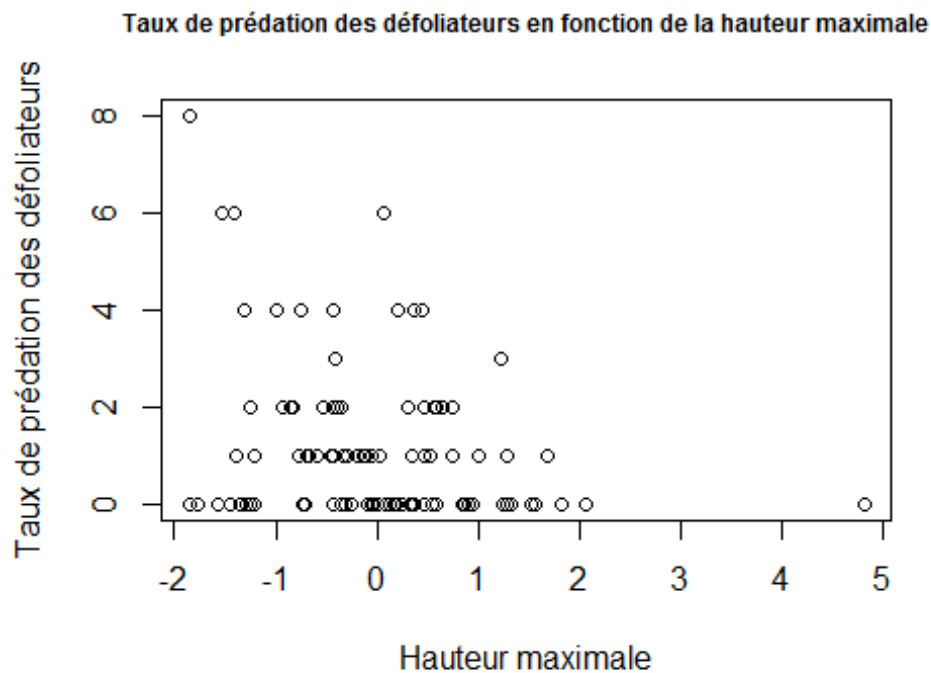
Boxplot de la proportion de 1er retours de faible intensité

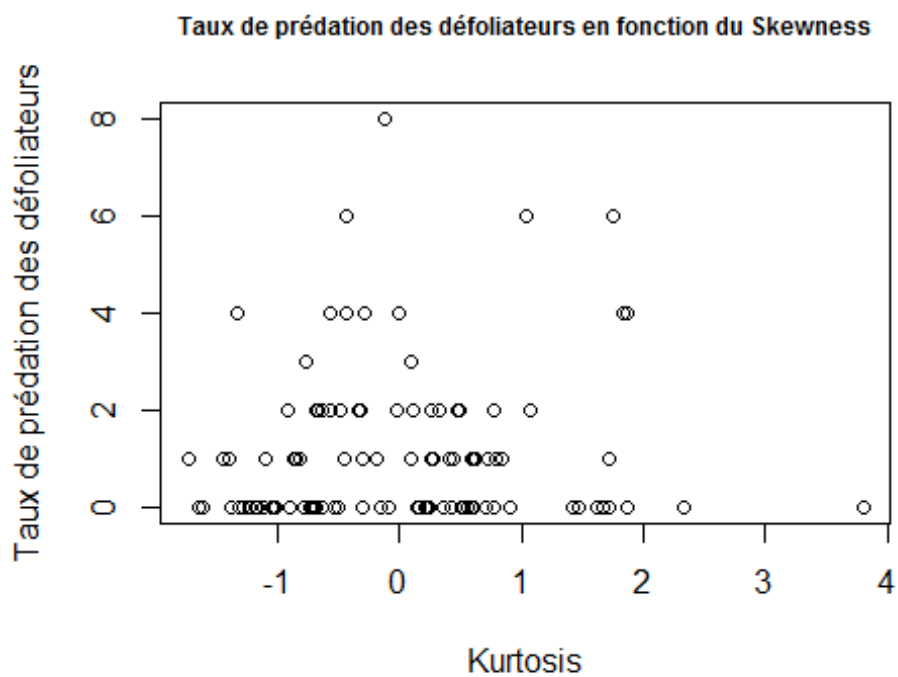
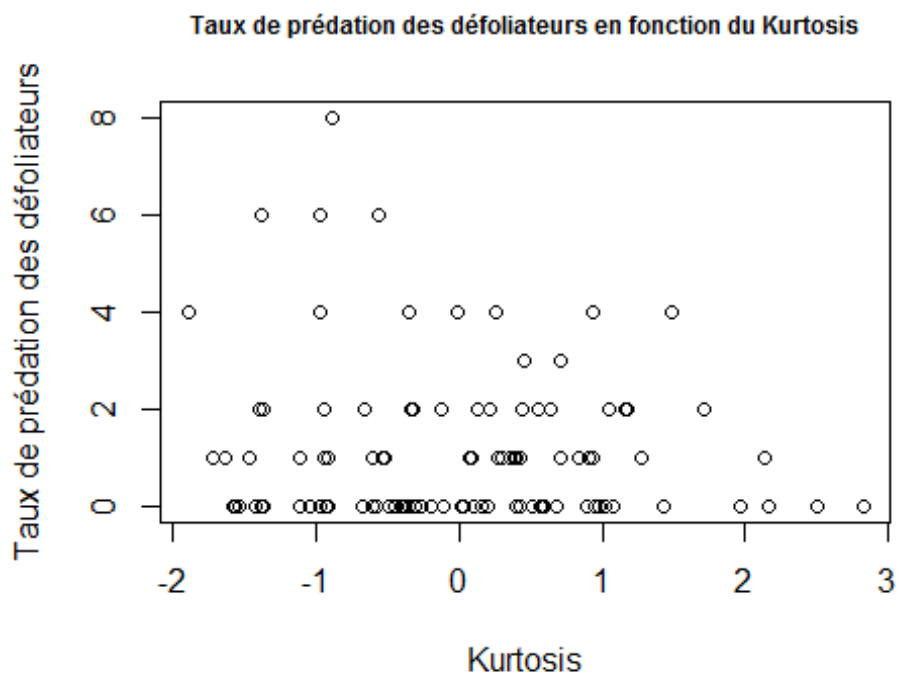


Boxplot du VCI

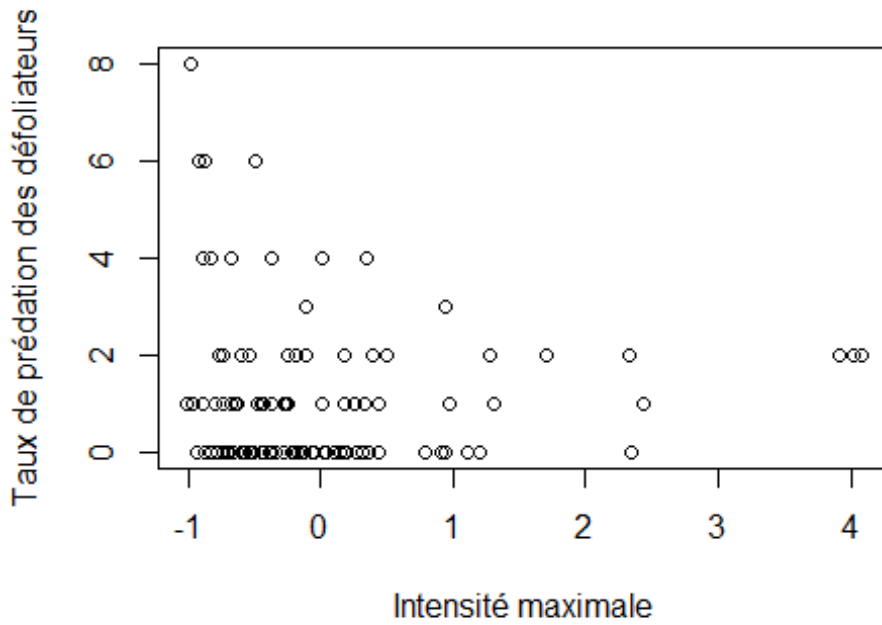


Diagrammes de dispersion du taux de prédation des chenilles en fonction des variables LiDAR non corrélées à l'échelle des plots de 26m :

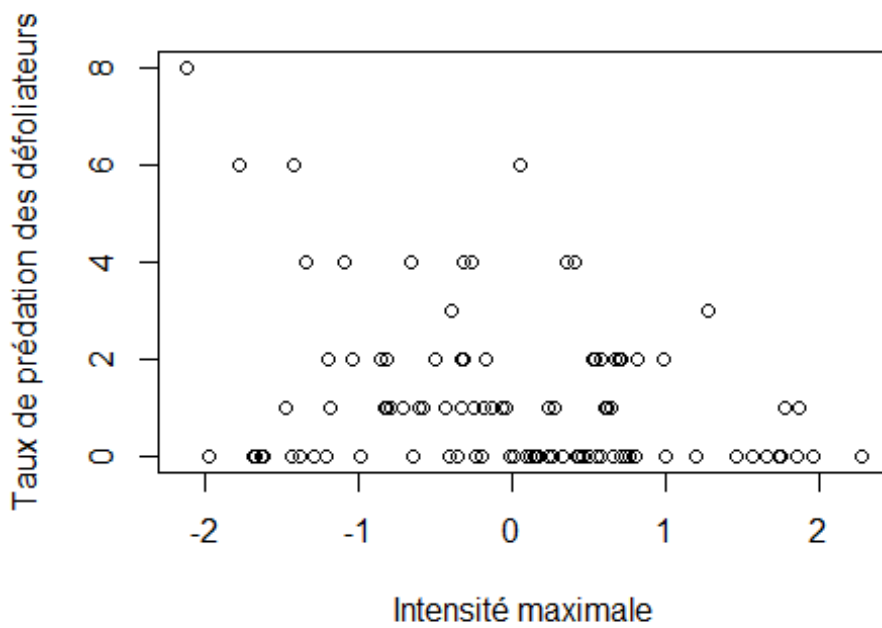




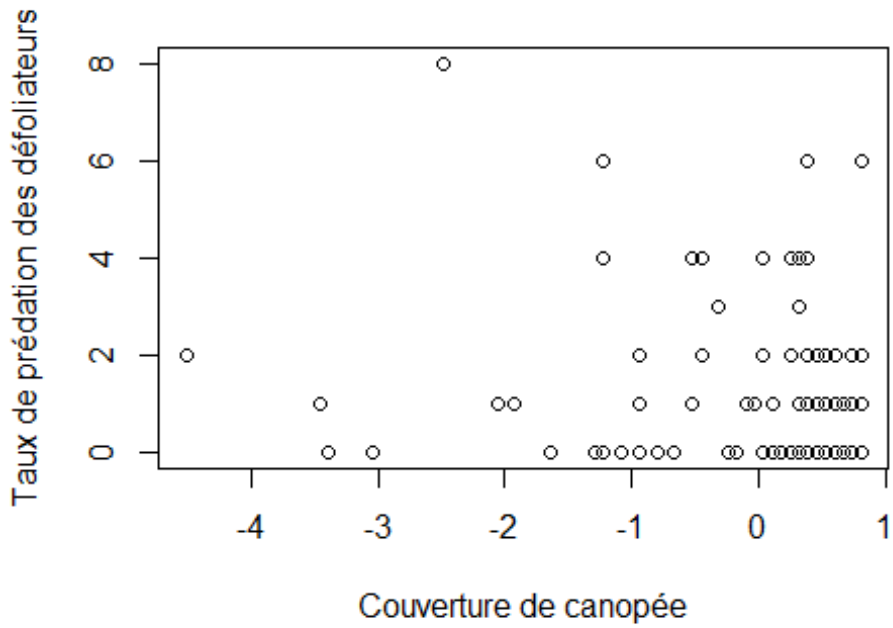
Taux de prédation des défoliateurs en fonction du percentile 1%



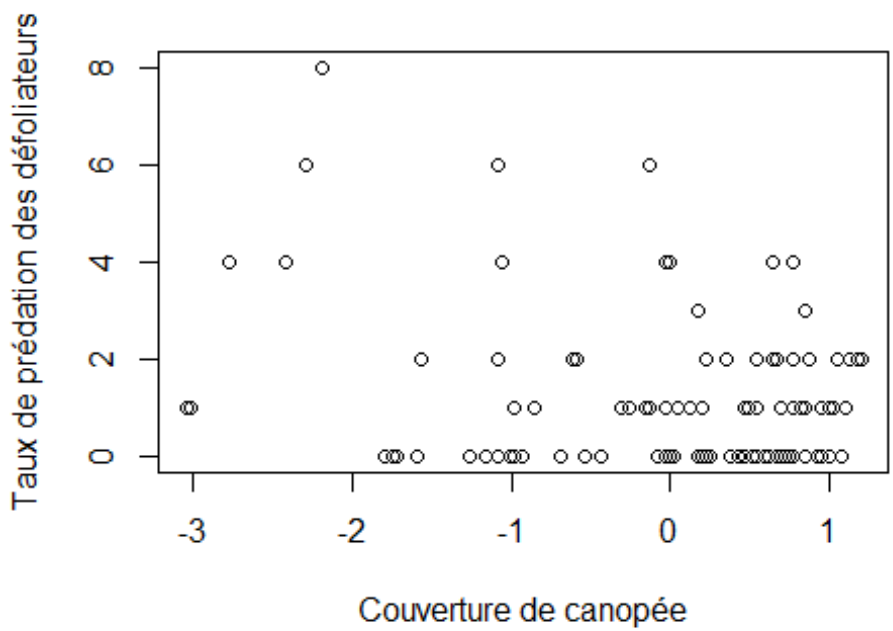
Taux de prédation des défoliateurs en fonction du percentile 1%



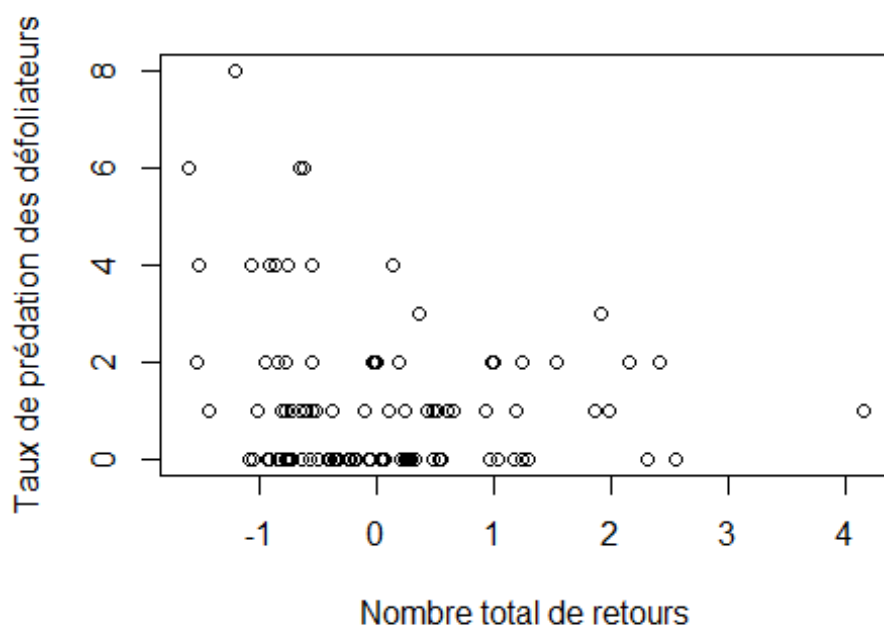
Taux de prédation des défoliateurs en fonction de la couverture de canopée



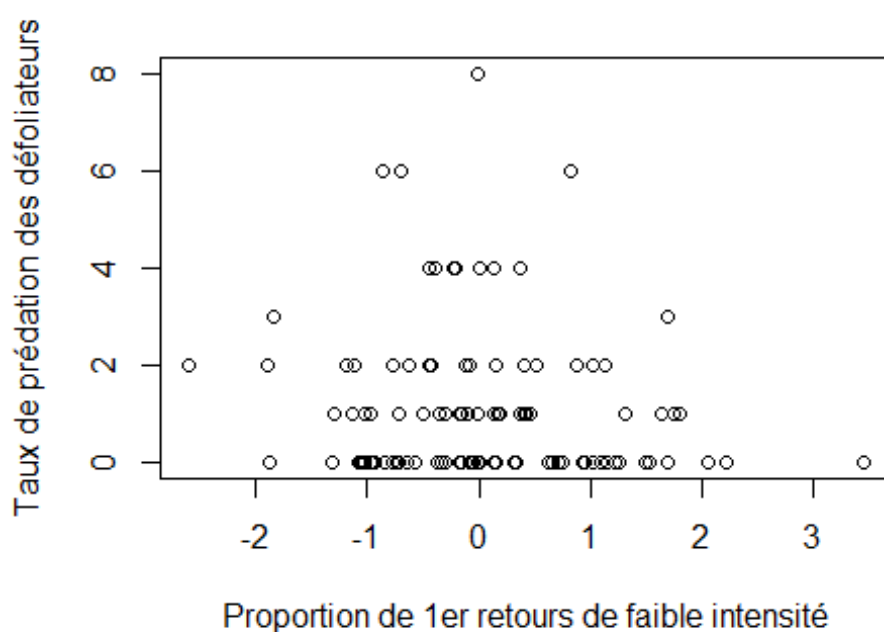
Taux de prédation des défoliateurs en fonction de la couverture de canopée à 71

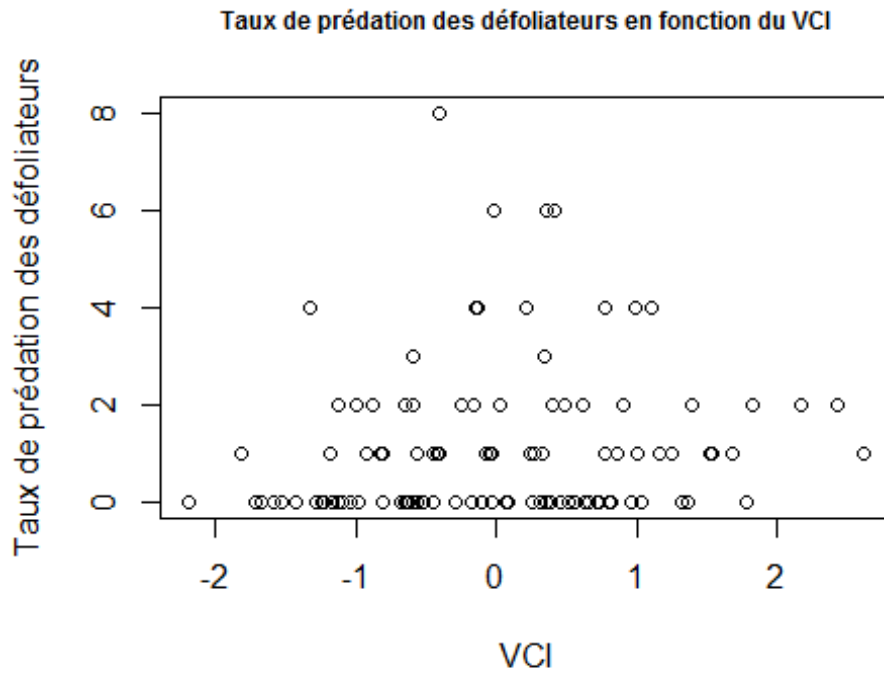


Taux de prédation des défoliateurs en fonction du nombre total de retours

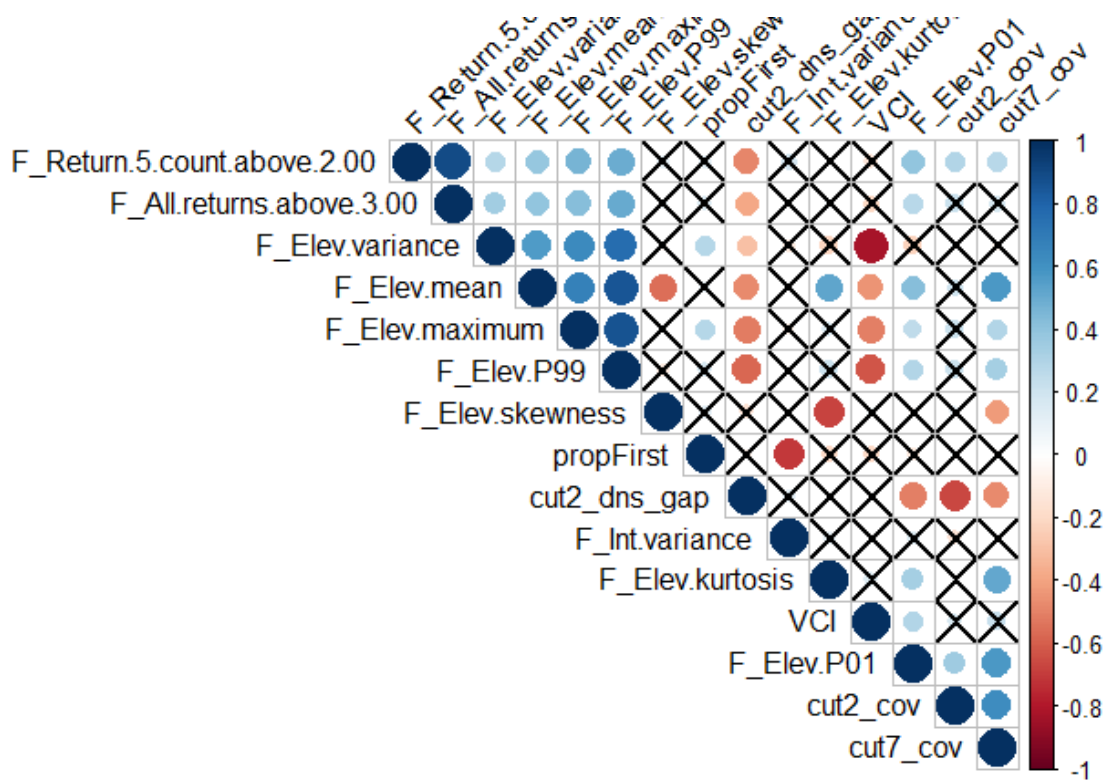


Taux de prédation des défoliateurs en fonction de la proportion de 1er retours de faible intensité





Corrélogramme des variables LiDAR peu corrélées à l'échelle des plots de 26m :



Résultats de la modélisation statistique par GLM modèle n°14

```
# Selection du meilleur modèle (variables significatives??)
mod_select =
glm(cbind(dataChenilleUseful$pred,dataChenilleUseful$tot_chenilles-
dataChenilleUseful$pred) ~
      cut2_cov + cut7_cov + F_Elev.mean + F_Elev.skewness +
F_Elev.P01,
      family = "binomial", data = dataMetricLidIBPMuesli)
summary(mod_select)

##
## Call:
## glm(formula = cbind(dataChenilleUseful$pred,
dataChenilleUseful$tot_chenilles -
##   dataChenilleUseful$pred) ~ cut2_cov + cut7_cov + F_Elev.mean +
##   F_Elev.skewness + F_Elev.P01, family = "binomial", data =
dataMetricLidIBPMuesli)
##
## Deviance Residuals:
##   Min       1Q   Median       3Q      Max
## -2.4684  -1.3159  -0.7929   0.5037   4.1265
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.83217    7.62535  -0.109  0.913098
## cut2_cov      0.15072    0.10247   1.471  0.141316
## cut7_cov     -0.16428    0.04687  -3.505  0.000457 ***
## F_Elev.mean  -0.24281    0.06590  -3.685  0.000229 ***
## F_Elev.skewness -1.40686    0.43835  -3.209  0.001330 **
## F_Elev.P01    0.70045    0.17385   4.029  5.6e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##   Null deviance: 247.77  on 103  degrees of freedom
## Residual deviance: 207.92  on  98  degrees of freedom
## AIC: 336.74
##
## Number of Fisher Scoring iterations: 5

# Analyse de La deviance du modele
D2 <- (mod_select$null.deviance - mod_select$deviance) /
mod_select$null.deviance
D2

## [1] 0.160855
```

Résultats de la modélisation statistique par GLMM modèle n°14

```

# Selection du meilleur modèle (AIC le plus faible)
mod_select_glmer = glmer(cbind(dataChenilleUseful_glmer$pred,
dataChenilleUseful_glmer$tot_chenilles -
                                dataChenilleUseful_glmer$pred) ~ cut7_cov +
F_Elev.mean +
                                F_Elev.P01 + F_Elev.skewness + (1 |
as.factor(dataChenilleUseful_glmer$bois)),
                                family = "binomial", data = dataMetricLidIBPMuesli,
na.action = "na.fail")
summary(mod_select_glmer)

## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula:
## cbind(dataChenilleUseful_glmer$pred, dataChenilleUseful_glmer$tot_chenilles
-
## dataChenilleUseful_glmer$pred) ~ cut7_cov + F_Elev.mean +
## F_Elev.P01 + F_Elev.skewness + (1 |
as.factor(dataChenilleUseful_glmer$bois))
## Data: dataMetricLidIBPMuesli
##
##      AIC      BIC  logLik deviance df.resid
##  308.0    323.9  -148.0   296.0     98
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.5352 -0.7364 -0.4371  0.6972  3.1281
##
## Random effects:
## Groups                Name                Variance Std.Dev.
## as.factor(dataChenilleUseful_glmer$bois) (Intercept) 0.8368   0.9148
## Number of obs: 104, groups: as.factor(dataChenilleUseful_glmer$bois), 28
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   11.17927   3.68721   3.032 0.002430 **
## cut7_cov      -0.13169   0.04365  -3.017 0.002554 **
## F_Elev.mean   -0.28997   0.08989  -3.226 0.001256 **
## F_Elev.P01     0.79904   0.21475   3.721 0.000199 ***
## F_Elev.skewness -1.22764   0.52046  -2.359 0.018336 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) ct7_cv F_Elv.m F_E.P0
## cut7_cov      -0.965
## F_Elev.mean   -0.086 -0.126
## F_Elev.P01     0.435 -0.453 -0.403
## F_Elv.skwns  -0.382  0.325  0.526 -0.406
## convergence code: 0

```

```
## unable to evaluate scaled gradient
## Model failed to converge: degenerate Hessian with 1 negative eigenvalues
r.squaredGLMM(mod_select_glmer)
## The result is correct only if all data used by the model has not changed
since model was fitted.
##           R2m           R2c
## 0.09713588 0.21113480
```

Résultats de la modélisation statistique par GLM modèle n°17

```
# Selection du meilleur modèle (AIC le plus faible)
mod_select = glm(dataRichesseUseful$Richesse.totale ~ F_Elev.maximum +
                 F_Elev.kurtosis + F_Int.maximum + F_Canopy.relief.ratio,
                 family = "gaussian", data = dataMetricLidScale)
summary(mod_select)

##
## Call:
## glm(formula = dataRichesseUseful$Richesse.totale ~ F_Elev.maximum +
##      F_Elev.kurtosis + F_Int.maximum + F_Canopy.relief.ratio,
##      family = "gaussian", data = dataMetricLidScale)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.08842  -0.39096  -0.06444   0.62110   2.88277
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      14.2500     0.2496  57.092 < 2e-16 ***
## F_Elev.maximum     0.8816     0.3006   2.933  0.00747 **
## F_Elev.kurtosis   -0.5902     0.2955  -1.997  0.05776 .
## F_Int.maximum     0.6137     0.2868   2.140  0.04322 *
## F_Canopy.relief.ratio 0.6197     0.3071   2.018  0.05541 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 1.744344)
##
##      Null deviance: 65.25  on 27  degrees of freedom
## Residual deviance: 40.12  on 23  degrees of freedom
## AIC: 101.53
##
## Number of Fisher Scoring iterations: 2

# Analyse de La deviance du modele
D2 <- (mod_select$null.deviance - mod_select$deviance) /
mod_select$null.deviance
D2

## [1] 0.3851354
```

Résultats de la modélisation statistique par GLM modèle n°18

```
# Selection du meilleur modèle (AIC Le plus faible)
mod_select = glm(dataRichesseUseful$Richesse.forestier ~ F_Elev.mean +
                 F_Elev.kurtosis + F_Int.maximum, family = "gaussian", data =
dataMetricLidScale)
summary(mod_select)

##
## Call:
## glm(formula = dataRichesseUseful$Richesse.forestier ~ F_Elev.mean +
##      F_Elev.kurtosis + F_Int.maximum, family = "gaussian", data =
dataMetricLidScale)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7230  -0.7364  -0.2073   0.6316   2.4559
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    6.5000     0.2292  28.354 <2e-16 ***
## F_Elev.mean     0.6419     0.2471   2.598  0.0158 *
## F_Elev.kurtosis -0.3985     0.2715  -1.468  0.1552
## F_Int.maximum   0.4583     0.2641   1.735  0.0956 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 1.471472)
##
##      Null deviance: 51.000  on 27  degrees of freedom
## Residual deviance: 35.315  on 24  degrees of freedom
## AIC: 95.96
##
## Number of Fisher Scoring iterations: 2

# Analyse de La deviance du modele
D2 <- (mod_select$null.deviance - mod_select$deviance) /
mod_select$null.deviance
D2

## [1] 0.3075427
```

Résultats de la modélisation statistique par GLM modèle n°3

```
# Selection du meilleur modèle (AIC Le plus faible)
mod_select = glm(dataIBPUseful$ibp ~ cut2_cov + F_Elev.mean + F_Elev.kurtosis +
                 F_Int.maximum + myTableScale$propFirst + myTableScale$VCI +
dataRichesseUseful$Nest.cavity +
as.numeric(dataNDVIUseful$NDVI_mean) +
                 as.numeric(dataNDVIUseful$NDVI_var),
                 family = "gaussian", data = dataMetricLidScale)
summary(mod_select)
```

```

##
## Call:
## glm(formula = dataIBPUseful$ibp ~ cut2_cov + F_Elev.mean + F_Elev.kurtosis +
##     F_Int.maximum + myTableScale$propFirst + myTableScale$VCI +
##     dataRichesseUseful$Nest.cavity + as.numeric(dataNDVIUseful$NDVI_mean) +
##     as.numeric(dataNDVIUseful$NDVI_var), family = "gaussian",
##     data = dataMetricLidScale)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -5.1305  -0.8082   0.1537   1.2935   5.4992
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    47.4578    20.3969   2.327  0.03185 *
## cut2_cov        1.4166     0.8008   1.769  0.09382 .
## F_Elev.mean    -1.3609     0.9501  -1.432  0.16917
## F_Elev.kurtosis  2.3150     1.0845   2.135  0.04679 *
## F_Int.maximum  -1.6244     0.9173  -1.771  0.09352 .
## myTableScale$propFirst  1.0005     0.7535   1.328  0.20079
## myTableScale$VCI    -3.7102     1.0480  -3.540  0.00234 **
## dataRichesseUseful$Nest.cavity  1.0062     0.5885   1.710  0.10451
## as.numeric(dataNDVIUseful$NDVI_mean) -32.0143    22.7469  -1.407  0.17634
## as.numeric(dataNDVIUseful$NDVI_var) 1011.6215   475.0404   2.130  0.04727 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 8.48626)
##
##     Null deviance: 362.96  on 27  degrees of freedom
## Residual deviance: 152.75  on 18  degrees of freedom
## AIC: 148.97
##
## Number of Fisher Scoring iterations: 2

# Analyse de la deviance du modele
D2 <- (mod_select$null.deviance - mod_select$deviance) /
mod_select$null.deviance
D2
## [1] 0.5791523

```

Résultats de la modélisation statistique par GLM modèle n°11

```

# Selection du meilleur modele (AIC Le plus faible)
mod_select =
glm(cbind(dataChenilleUseful$pred,dataChenilleUseful$tot_chenilles-
dataChenilleUseful$pred) ~
      F_Elev.mean + F_Int.maximum +
      F_Total.all.returns + F_Canopy.relief.ratio +
myTableScale$VCI +
      as.numeric(dataNDVIUseful$NDVI_mean) +

```

```

dataRichesseUseful$Insect,
      family = "binomial", data = dataMetricLidScale)
summary(mod_select)

##
## Call:
## glm(formula = cbind(dataChenilleUseful$pred,
dataChenilleUseful$tot_chenilles ~
##   dataChenilleUseful$pred) ~ F_Elev.mean + F_Int.maximum +
##   F_Total.all.returns + F_Canopy.relief.ratio + myTableScale$VCI +
##   as.numeric(dataNDVIUseful$NDVI_mean) + dataRichesseUseful$Insect,
##   family = "binomial", data = dataMetricLidScale)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.20097  -1.02236   0.04501   0.93549   1.89962
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      7.54063    2.90674   2.594 0.00948
## F_Elev.mean     -0.82696    0.16123  -5.129 2.91e-07
## F_Int.maximum    0.38712    0.11927   3.246 0.00117
## F_Total.all.returns 0.38860    0.14181   2.740 0.00614
## F_Canopy.relief.ratio 0.64688    0.16271   3.976 7.02e-05
## myTableScale$VCI -0.73608    0.15796  -4.660 3.16e-06
## as.numeric(dataNDVIUseful$NDVI_mean) -7.25453    3.17700  -2.283 0.02240
## dataRichesseUseful$Insect -0.43607    0.08687  -5.020 5.17e-07
##
##              **
## F_Elev.mean   ***
## F_Int.maximum **
## F_Total.all.returns **
## F_Canopy.relief.ratio ***
## myTableScale$VCI ***
## as.numeric(dataNDVIUseful$NDVI_mean) *
## dataRichesseUseful$Insect ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 110.807  on 27  degrees of freedom
## Residual deviance:  40.705  on 20  degrees of freedom
## AIC: 131.38
##
## Number of Fisher Scoring iterations: 5

# Analyse de La deviance du modele
D2 <- (mod_select>null.deviance - mod_select$deviance) /
mod_select>null.deviance
D2

```



```
## [1] 0.6326456
```

Résultats de la modélisation statistique par GLM modèle n°10

```
# Selection du meilleur modèle (AIC Le plus faible)
```

```
mod_select =  
glm(cbind(dataChenilleUseful$pred,dataChenilleUseful$tot_chenilles -  
dataChenilleUseful$pred) ~  
      F_Elev.mean + F_Elev.kurtosis +  
      F_Int.maximum + cut2_dns_gap + F_Total.all.returns +  
F_Canopy.relief.ratio +  
      myTableScale$propFirst + myTableScale$VCI +  
dataAbondanceUseful$Insect,  
      family = "binomial", data = dataMetricLidScale)  
summary(mod_select)  
  
##  
## Call:  
## glm(formula = cbind(dataChenilleUseful$pred,  
dataChenilleUseful$tot_chenilles -  
##   dataChenilleUseful$pred) ~ F_Elev.mean + F_Elev.kurtosis +  
##   F_Int.maximum + cut2_dns_gap + F_Total.all.returns +  
F_Canopy.relief.ratio +  
##   myTableScale$propFirst + myTableScale$VCI + dataAbondanceUseful$Insect,  
##   family = "binomial", data = dataMetricLidScale)  
##  
## Deviance Residuals:  
##   Min       1Q   Median       3Q      Max  
## -3.02790 -0.77530  0.05803  0.61713  2.32211  
##  
## Coefficients:  
##              Estimate Std. Error z value Pr(>|z|)  
## (Intercept)      1.46840    0.77736   1.889 0.058896 .  
## F_Elev.mean     -1.48359    0.25651  -5.784 7.31e-09 ***  
## F_Elev.kurtosis  0.35894    0.16489   2.177 0.029495 *  
## F_Int.maximum    0.33483    0.16199   2.067 0.038737 *  
## cut2_dns_gap    -0.61746    0.18046  -3.422 0.000622 ***  
## F_Total.all.returns 0.43344    0.16638   2.605 0.009184 **  
## F_Canopy.relief.ratio 1.08193    0.22093   4.897 9.73e-07 ***  
## myTableScale$propFirst -0.34924    0.15289  -2.284 0.022353 *  
## myTableScale$VCI   -1.24265    0.24485  -5.075 3.87e-07 ***  
## dataAbondanceUseful$Insect -0.27727    0.06124  -4.527 5.98e-06 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## (Dispersion parameter for binomial family taken to be 1)  
##  
##   Null deviance: 110.807  on 27  degrees of freedom  
## Residual deviance:  41.782  on 18  degrees of freedom  
## AIC: 136.45  
##  
## Number of Fisher Scoring iterations: 5
```

```
# Analyse de la deviance du modele
```

```
D2 <- (mod_select$null.deviance - mod_select$deviance) /  
mod_select$null.deviance  
D2
```

```
## [1] 0.6229248
```

Résultats de la modélisation statistique par GLM modèle n°13

```
# Selection du meilleur modele (AIC Le plus faible)
```

```
mod_select =  
glm(cbind(dataChenilleUseful$pred,dataChenilleUseful$tot_chenilles -  
dataChenilleUseful$pred) ~  
      F_Elev.mean + F_Elev.kurtosis +  
      F_Total.all.returns + F_Canopy.relief.ratio +  
myTableScale$VCI +  
      as.numeric(dataNDVIUseful$NDVI_mean) +  
as.numeric(dataNDVIUseful$NDVI_var) +  
      dataRichesseUseful$Forage.canopy,  
      family = "binomial", data = dataMetricLidScale)  
summary(mod_select)  
  
##  
## Call:  
## glm(formula = cbind(dataChenilleUseful$pred,  
dataChenilleUseful$tot_chenilles -  
##   dataChenilleUseful$pred) ~ F_Elev.mean + F_Elev.kurtosis +  
##   F_Total.all.returns + F_Canopy.relief.ratio + myTableScale$VCI +  
##   as.numeric(dataNDVIUseful$NDVI_mean) +  
as.numeric(dataNDVIUseful$NDVI_var) +  
##   dataRichesseUseful$Forage.canopy, family = "binomial", data =  
dataMetricLidScale)  
##  
## Deviance Residuals:  
##      Min       1Q   Median       3Q      Max   
## -2.48230 -1.02442 -0.00825  0.29596  3.08518   
##  
## Coefficients:  
##              Estimate Std. Error z value Pr(>|z|)   
## (Intercept)          5.3934     2.9949   1.801  0.07172   
## F_Elev.mean         -1.0947     0.2097  -5.221 1.78e-07   
## F_Elev.kurtosis      0.4657     0.1503   3.099 0.00194   
## F_Total.all.returns  0.3761     0.1360   2.767 0.00566   
## F_Canopy.relief.ratio 0.5744     0.1886   3.046 0.00232   
## myTableScale$VCI    -0.8753     0.1818  -4.814 1.48e-06   
## as.numeric(dataNDVIUseful$NDVI_mean) -5.9797     3.2847  -1.821 0.06868   
## as.numeric(dataNDVIUseful$NDVI_var) 116.1424    69.1149   1.680 0.09287   
## dataRichesseUseful$Forage.canopy    -0.5113     0.1693  -3.020 0.00253   
##  
## (Intercept)          .   
## F_Elev.mean          ***   
## F_Elev.kurtosis      **
```

```

## F_Total.all.returns          **
## F_Canopy.relief.ratio        **
## myTableScale$VCI             ***
## as.numeric(dataNDVIUseful$NDVI_mean) .
## as.numeric(dataNDVIUseful$NDVI_var) .
## dataRichesseUseful$Forage.canopy **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 110.807  on 27  degrees of freedom
## Residual deviance:  48.541  on 19  degrees of freedom
## AIC: 141.21
##
## Number of Fisher Scoring iterations: 5

# Analyse de la deviance du modele
D2 <- (mod_select>null.deviance - mod_select$deviance) /
mod_select>null.deviance
D2

## [1] 0.5619321

```

Résultats de la modélisation statistique par GLM modèle n°12

```

# Selection du meilleur modele (AIC Le plus faible)
mod_select =
glm(cbind(dataChenilleUseful$pred,dataChenilleUseful$tot_chenilles-
dataChenilleUseful$pred) ~
      F_Elev.mean + F_Elev.kurtosis +
      cut2_dns_gap + F_Total.all.returns + F_Canopy.relief.ratio +
      myTableScale$propFirst + myTableScale$VCI +
dataAbondanceUseful$Forage.canopy,
      family = "binomial", data = dataMetricLidScale)
summary(mod_select)

##
## Call:
## glm(formula = cbind(dataChenilleUseful$pred,
dataChenilleUseful$tot_chenilles -
##      dataChenilleUseful$pred) ~ F_Elev.mean + F_Elev.kurtosis +
##      cut2_dns_gap + F_Total.all.returns + F_Canopy.relief.ratio +
##      myTableScale$propFirst + myTableScale$VCI +
dataAbondanceUseful$Forage.canopy,
##      family = "binomial", data = dataMetricLidScale)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.7484  -1.1694  -0.2083   0.6996   2.8047
##
## Coefficients:

```

```

##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      0.82441    0.74865   1.101 0.270814
## F_Elev.mean     -1.67111    0.27813  -6.008 1.87e-09 ***
## F_Elev.kurtosis  0.75086    0.18968   3.959 7.54e-05 ***
## cut2_dns_gap    -0.49695    0.17015  -2.921 0.003492 **
## F_Total.all.returns 0.23461    0.13190   1.779 0.075293 .
## F_Canopy.relief.ratio 0.86836    0.20393   4.258 2.06e-05 ***
## myTableScale$propFirst -0.24201    0.12185  -1.986 0.047012 *
## myTableScale$VCI    -1.06720    0.19950  -5.349 8.82e-08 ***
## dataAbondanceUseful$Forage.canopy -0.33752    0.08762  -3.852 0.000117 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 110.807 on 27 degrees of freedom
## Residual deviance: 50.219 on 19 degrees of freedom
## AIC: 142.89
##
## Number of Fisher Scoring iterations: 5

# Analyse de la deviance du modele
D2 <- (mod_select$null.deviance - mod_select$deviance) /
mod_select$null.deviance
D2

## [1] 0.5467912

```

Résultats de la modélisation statistique par GLM modèle n°19

```

# Selection du meilleur modele (AIC Le plus faible)
mod_select = glm(dataRichesseUseful$Nest.cavity ~ F_Elev.maximum +
                F_Elev.mean + F_Elev.kurtosis + F_Int.maximum +
                F_Canopy.relief.ratio +
                myTableScale$VCI, family = "gaussian", data =
                dataMetricLidScale)
summary(mod_select)

##
## Call:
## glm(formula = dataRichesseUseful$Nest.cavity ~ F_Elev.maximum +
##      F_Elev.mean + F_Elev.kurtosis + F_Int.maximum + F_Canopy.relief.ratio +
##      myTableScale$VCI, family = "gaussian", data = dataMetricLidScale)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.98087  -0.42755  -0.03595   0.63204   1.77696
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      4.1429    0.2025  20.461 2.38e-15 ***
## F_Elev.maximum   -2.0726    1.2397  -1.672  0.10938

```

```

## F_Elev.mean          2.9416      1.3595      2.164      0.04216 *
## F_Elev.kurtosis      -1.2546      0.3602     -3.484      0.00222 **
## F_Int.maximum        0.8091      0.2453      3.298      0.00343 **
## F_Canopy.relief.ratio -1.9927      1.2115     -1.645      0.11489
## myTableScale$VCI     0.8465      0.3836      2.207      0.03860 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 1.147953)
##
## Null deviance: 47.429  on 27  degrees of freedom
## Residual deviance: 24.107  on 21  degrees of freedom
## AIC: 91.269
##
## Number of Fisher Scoring iterations: 2

# Analyse de la deviance du modele
D2 <- (mod_select>null.deviance - mod_select$deviance) /
mod_select>null.deviance
D2

## [1] 0.4917196

```

Résultats de la modélisation statistique par GLM modèle n°21

```

mod_select = glm(dataRichesseUseful$Nest.canopy ~ F_Elev.maximum +
                F_Elev.mean + F_Canopy.relief.ratio + myTableScale$VCI,
                family = "gaussian", data = dataMetricLidScale)
summary(mod_select)

##
## Call:
## glm(formula = dataRichesseUseful$Nest.canopy ~ F_Elev.maximum +
##      F_Elev.mean + F_Canopy.relief.ratio + myTableScale$VCI, family =
##      "gaussian",
##      data = dataMetricLidScale)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8582  -0.7502  -0.1306   0.7908   1.6259
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      4.3571     0.2068  21.069 <2e-16 ***
## F_Elev.maximum    2.0955     1.0904   1.922  0.0671 .
## F_Elev.mean      -2.0190     1.0996  -1.836  0.0793 .
## F_Canopy.relief.ratio 1.6558     1.0508   1.576  0.1288
## myTableScale$VCI  -0.5059     0.2615  -1.935  0.0654 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 1.197457)

```

```
##
## Null deviance: 38.429 on 27 degrees of freedom
## Residual deviance: 27.542 on 23 degrees of freedom
## AIC: 90.998
##
## Number of Fisher Scoring iterations: 2

# Analyse de La deviance du modele
D2 <- (mod_select$null.deviance - mod_select$deviance) /
mod_select$null.deviance
D2

## [1] 0.2833064
```

Résultats de la modélisation statistique par GLM modèle n°26

```
mod_select = glm(dataRichesseUseful$Insect ~ F_Elev.mean + F_Elev.kurtosis,
                 family = "gaussian", data = dataMetricLidScale)
summary(mod_select)

##
## Call:
## glm(formula = dataRichesseUseful$Insect ~ F_Elev.mean + F_Elev.kurtosis,
##      family = "gaussian", data = dataMetricLidScale)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.2291  -0.7275   0.1959   0.7425   2.3260
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      7.4286     0.2338  31.776 < 2e-16 ***
## F_Elev.mean       0.7406     0.2509   2.951  0.00679 **
## F_Elev.kurtosis  -0.4513     0.2509  -1.798  0.08419 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 1.53028)
##
## Null deviance: 52.857 on 27 degrees of freedom
## Residual deviance: 38.257 on 25 degrees of freedom
## AIC: 96.2
##
## Number of Fisher Scoring iterations: 2

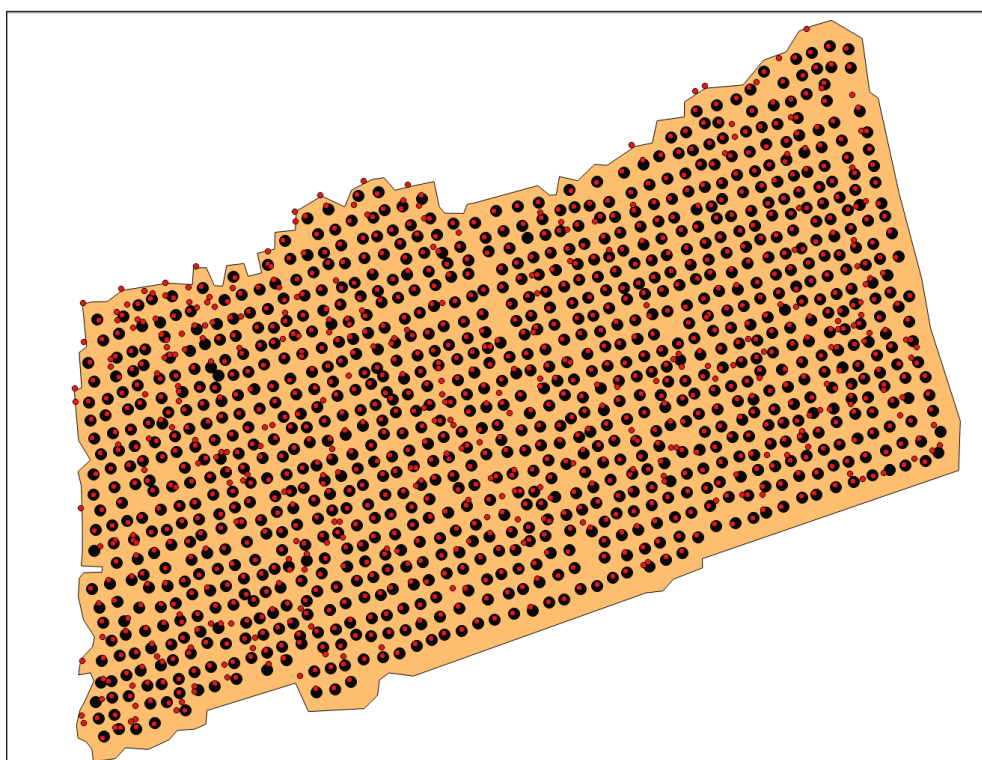
# Analyse de La deviance du modele
D2 <- (mod_select$null.deviance - mod_select$deviance) /
mod_select$null.deviance
D2

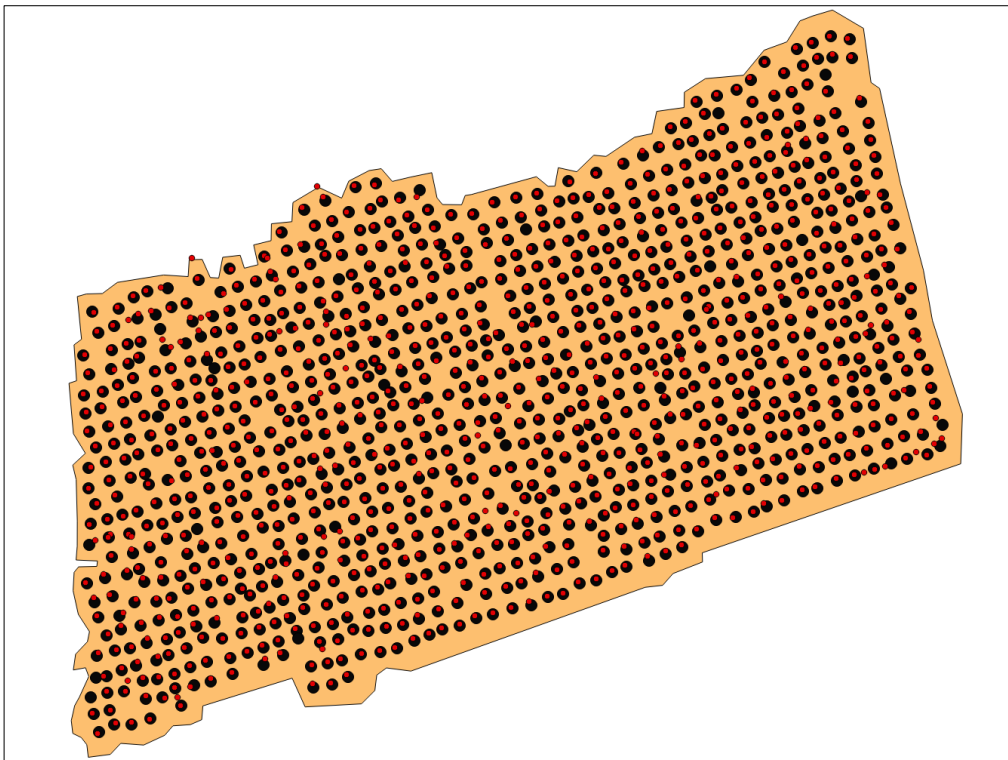
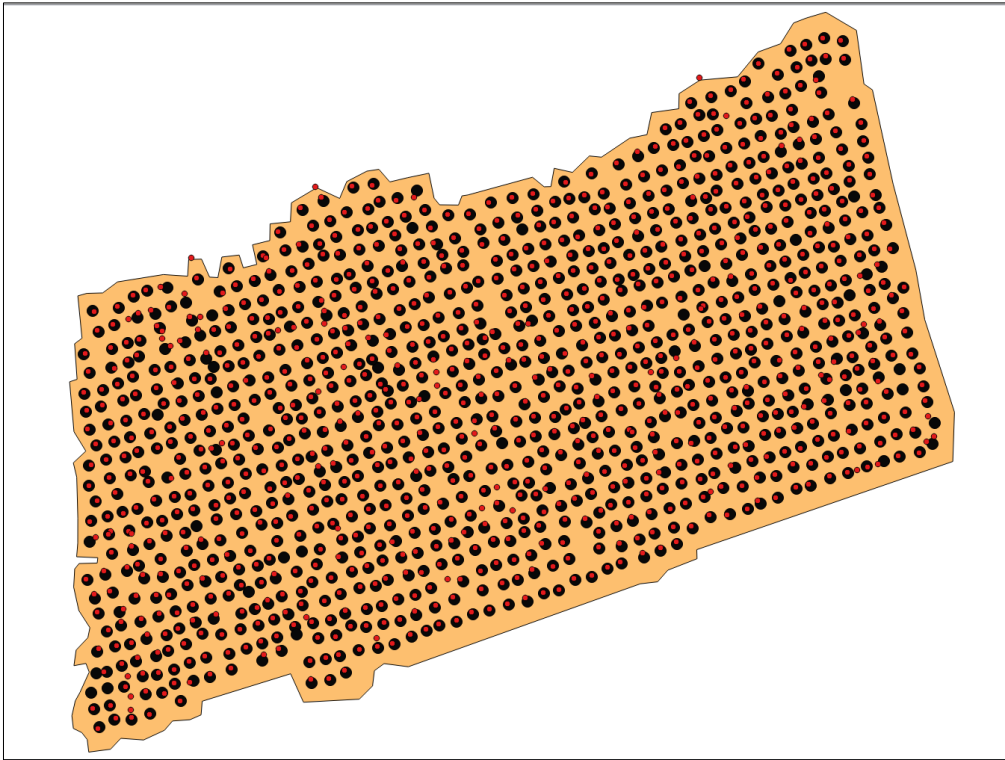
## [1] 0.276219
```

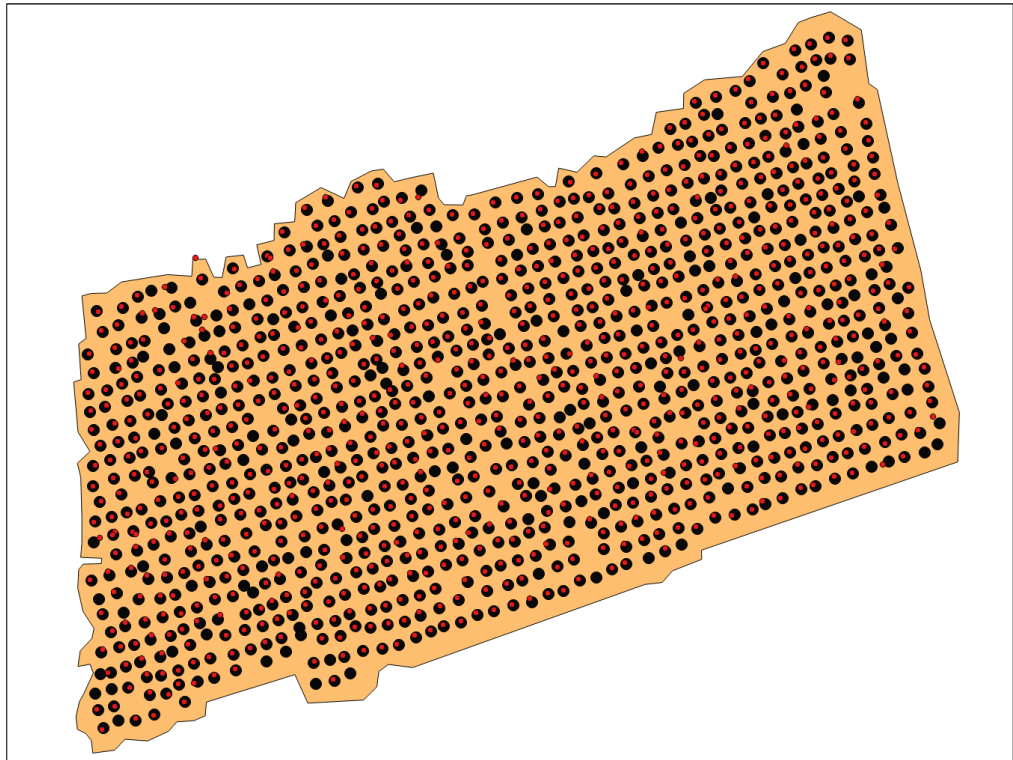
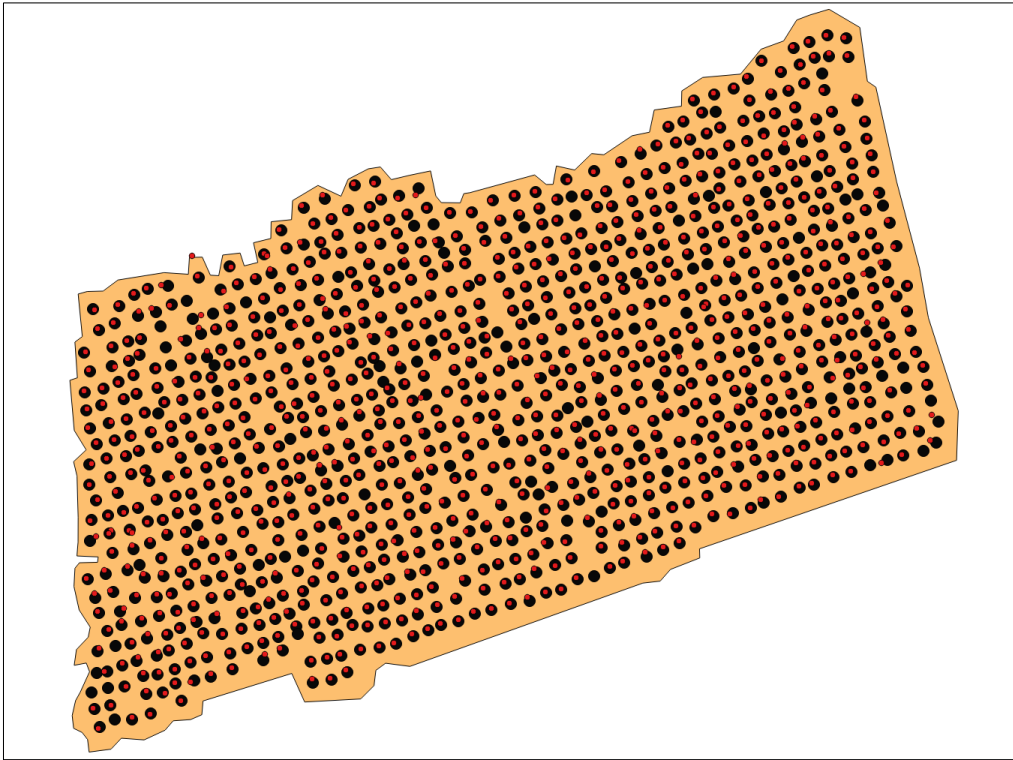
Annexe 6 : Résultats de la segmentation

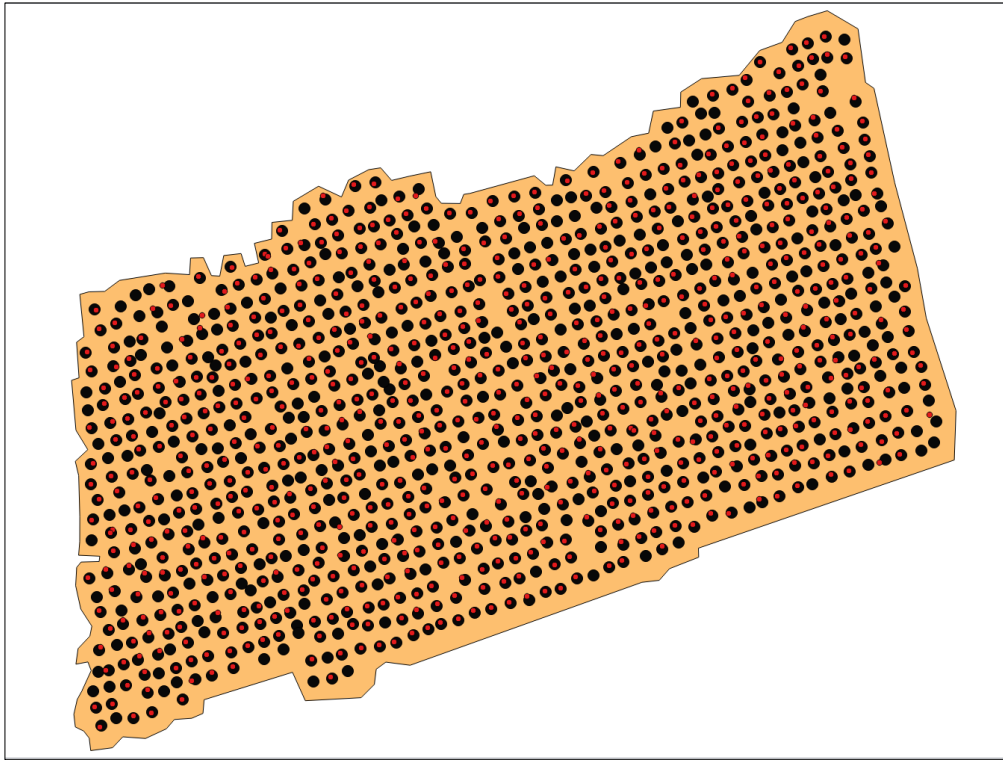
Résultats avec l'algorithme Python :

Sans Filtrage - Rayon de recherche = 2.5m/3m/3.5m/4m/4.5m/5m (de haut en bas) :

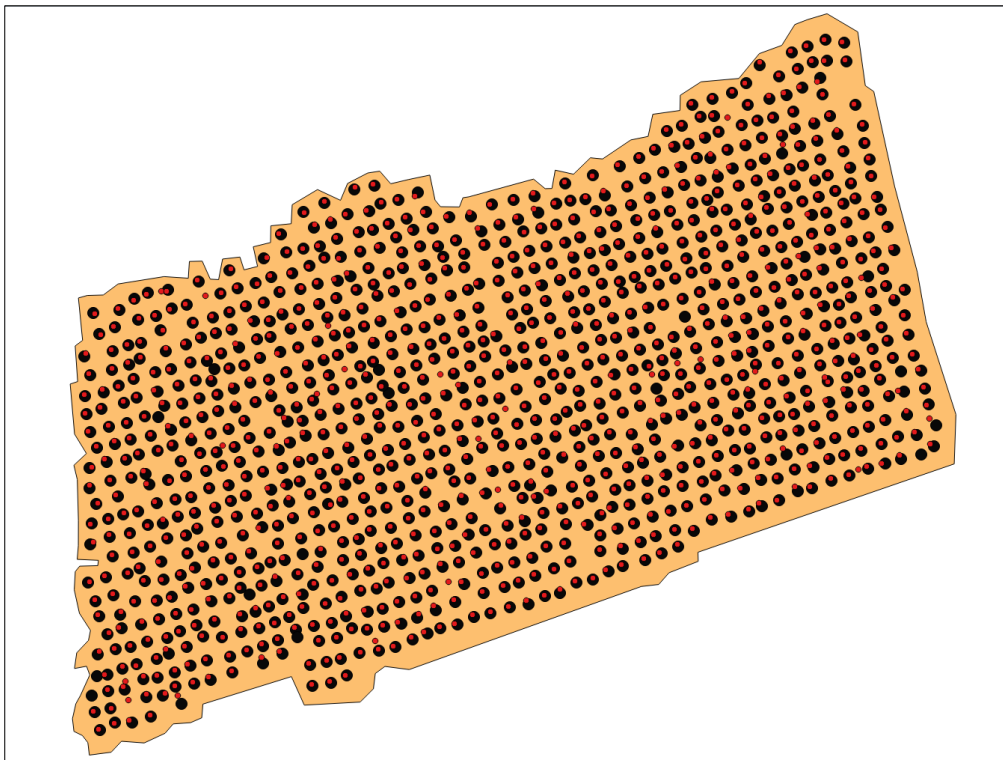


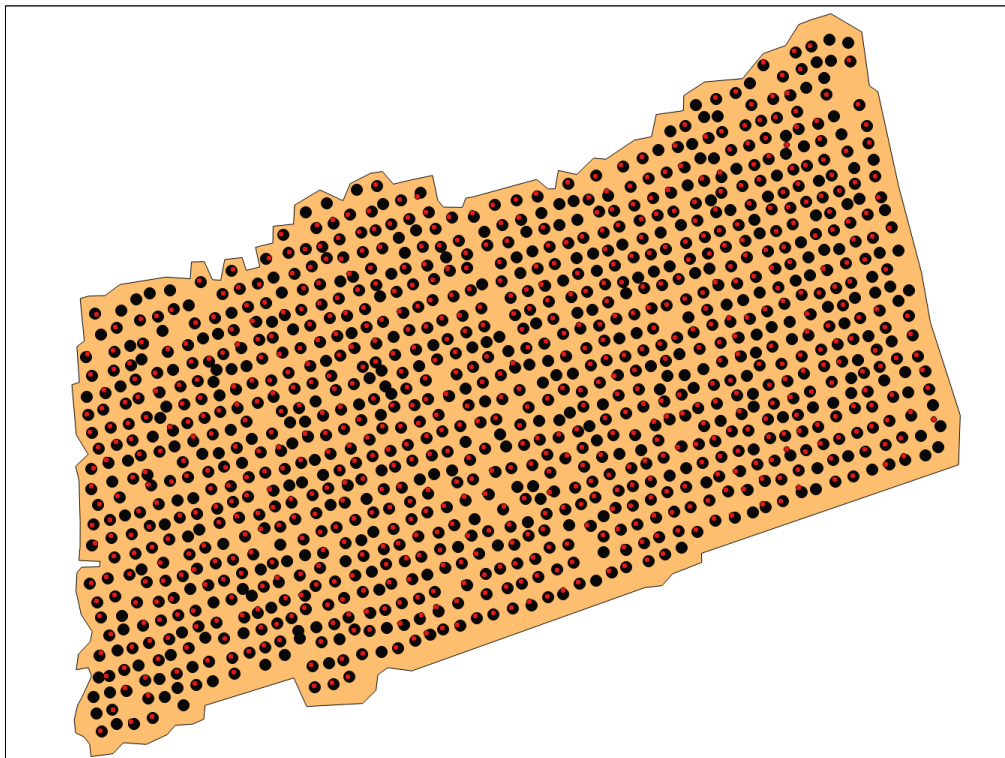
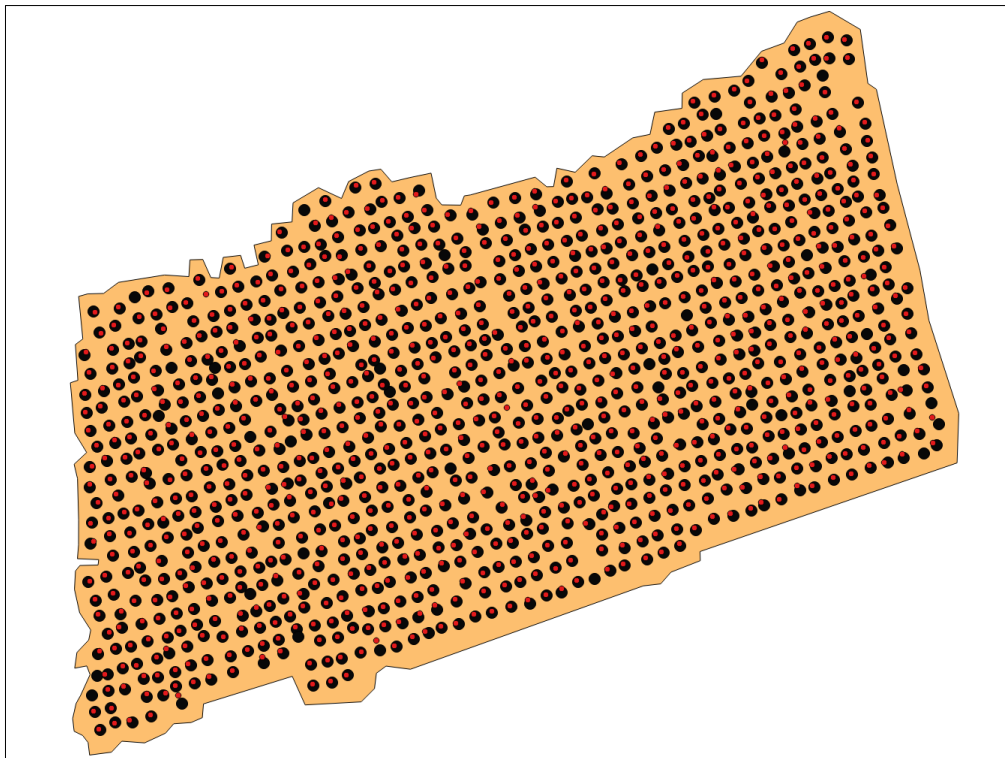




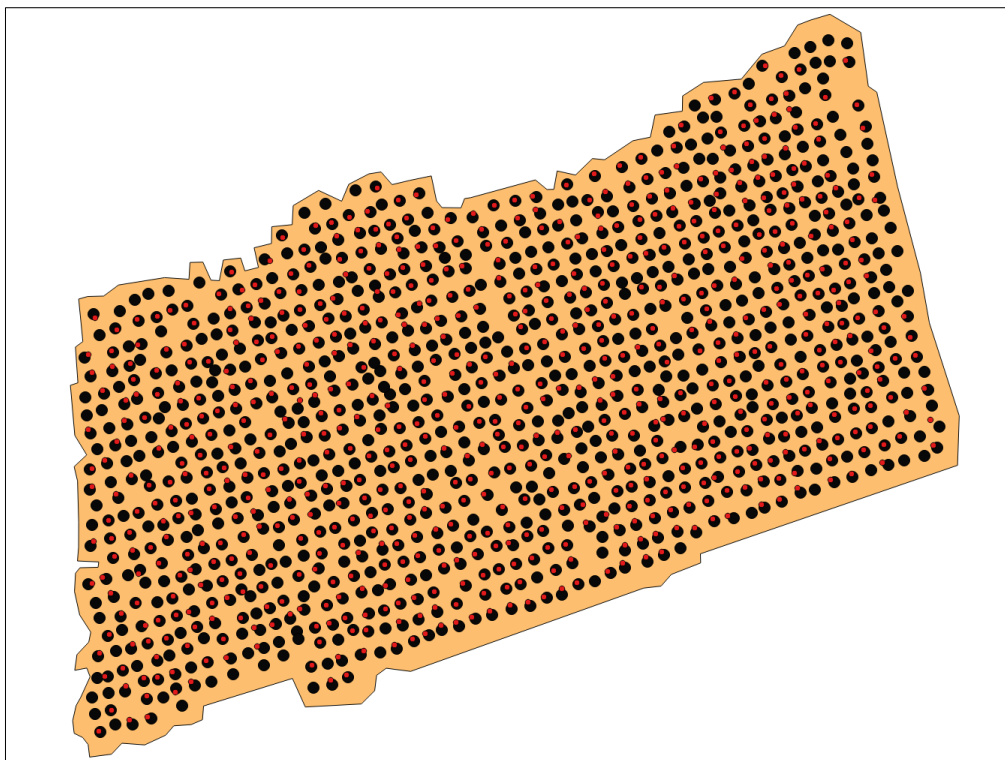
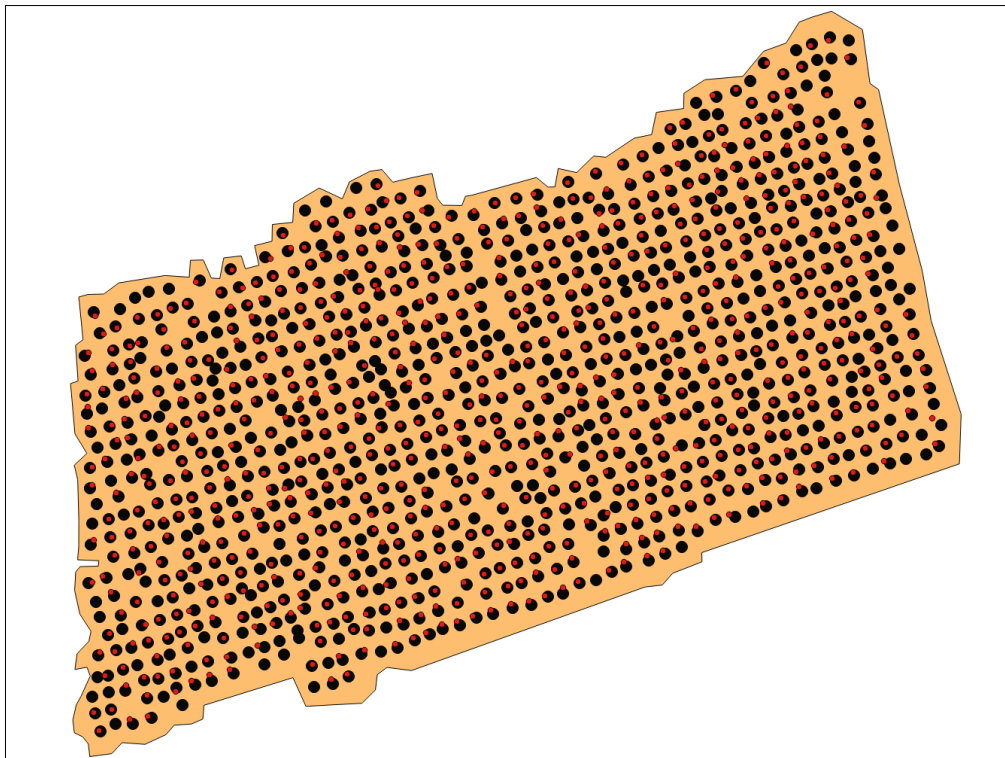


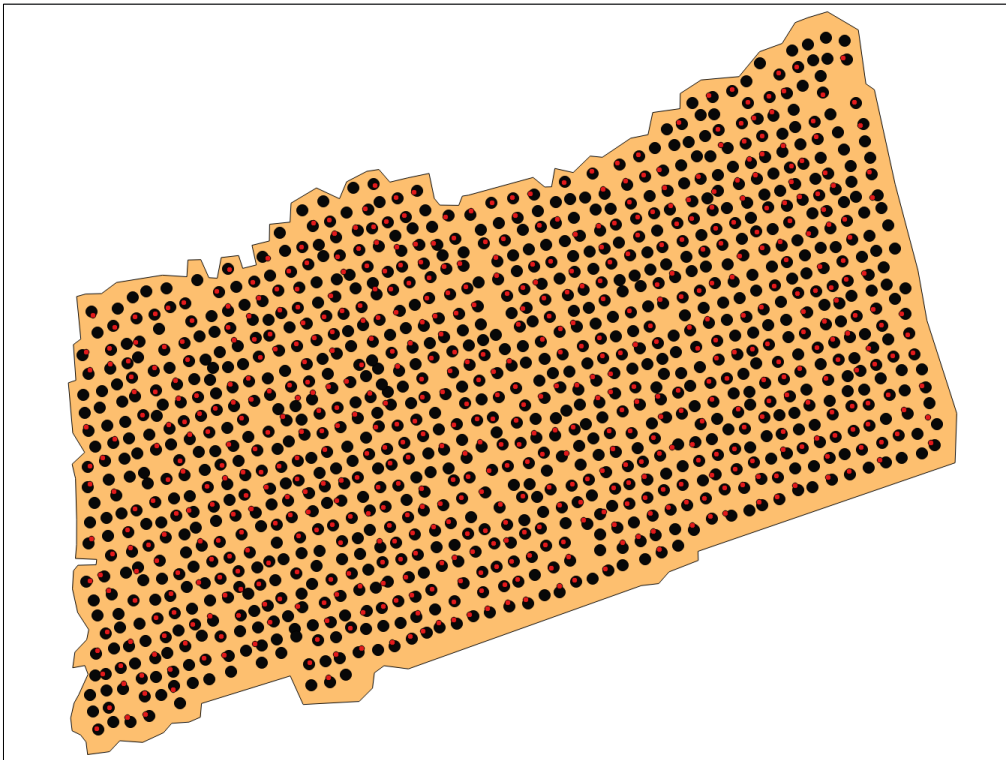
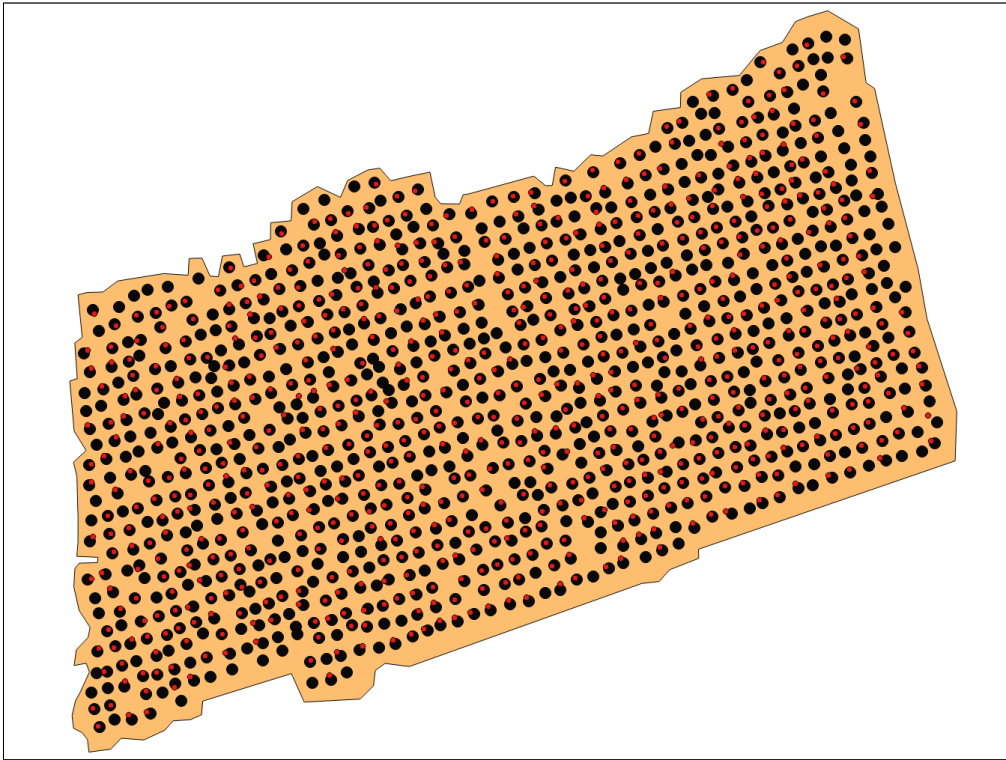
Filtrage R1 - Rayon de recherche = 2.5m/3m/5m (de haut en bas) :

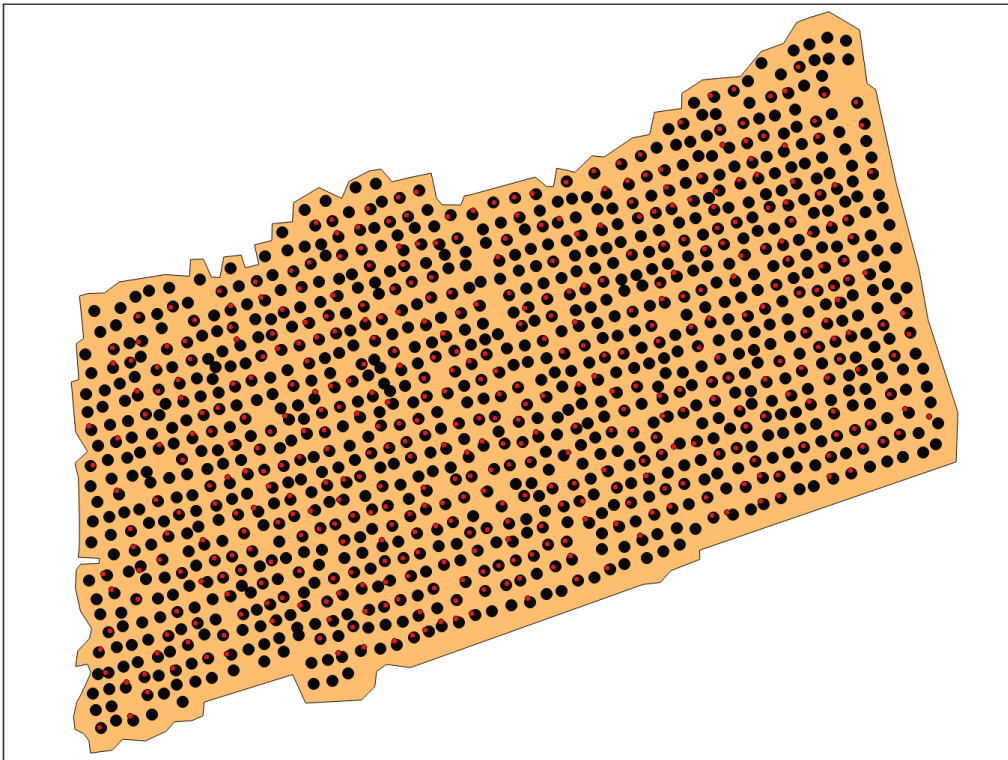
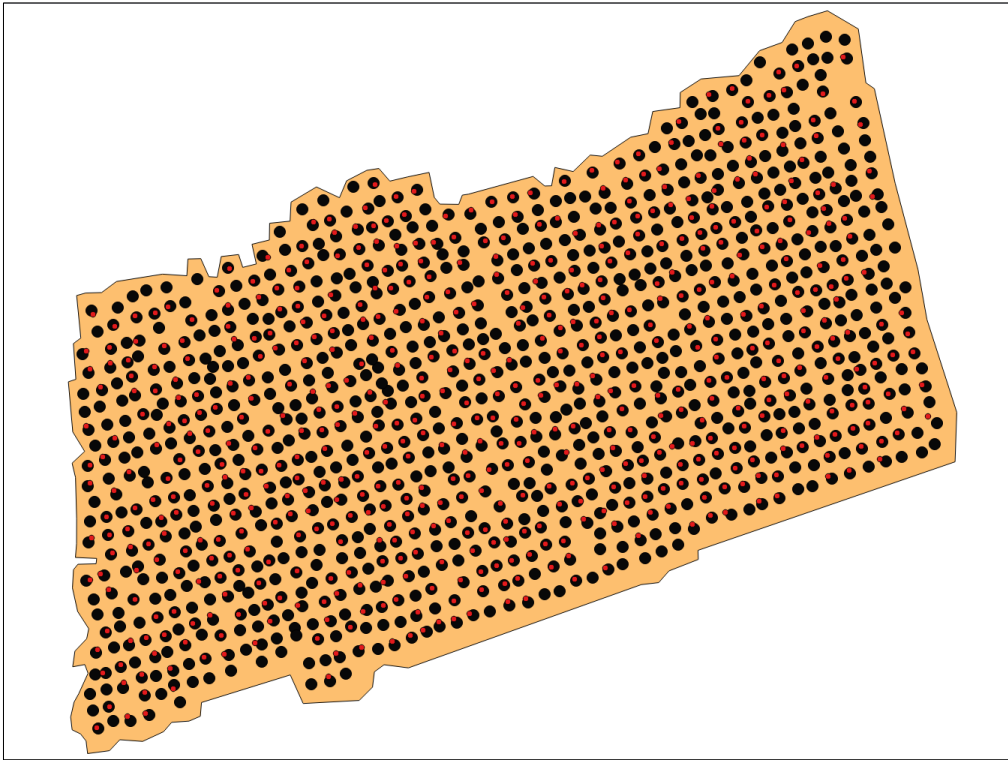




Filtrage R2 - Rayon de recherche = 2.5m/3m/3.5m/4m/4.5m/5m (de haut en bas) :







Annexe 7 : Modélisation statistique par PLS

L'analyse par régression PLS est une méthode itérative développée par Herman Wold. Elle combine à la fois la démarche de la régression et celle de l'Analyse en Composantes Principales (ACP). En effet, nous obtenons des composantes qui sont des combinaisons linéaires des variables « X » et qui jouent le rôle des variables explicatives. Et à partir de ces nouveaux facteurs un modèle de régression sur les variables réponses va être construit, prenant en compte les corrélations entre les variables réponses et explicatives.

Nous avons choisi d'appliquer cette méthode exploratoire pour deux raisons principales. D'une part, elle permet de s'affranchir de l'utilisation des seules variables peu corrélées entre elles. Nous effectuons donc cette modélisation à l'aide de l'ensemble des métriques LiDAR extraites du nuage de points (une centaine). D'autre part, cette méthode ne nous oblige plus à nous restreindre à la construction de modèles expliquant ou prédisant une seule variable réponse. En effet, plusieurs « Y » peuvent être utilisés. Cela permet ainsi de trouver des liens entre plusieurs indicateurs de biodiversité à partir de nombreuses variables LiDAR potentiellement corrélées entre elles. Binard (2012) définit d'ailleurs la régression PLS comme une méthode qui « permet la construction de modèles prédictifs quand les variables sont nombreuses et fortement corrélées entre elles. Cette méthode peut également être utilisée quand le nombre de variables dépasse celui des observations ».

Pour tester cette méthode, nous avons décidé de construire le modèle tableau 12 incluant plusieurs X et Y.

Tableau 12 Modèle PLS calculé.

Modèle créé à l'échelle des bois MJESLI	
Variables réponses	Variables explicatives
IBP total, indice de consommation lié au chevreuil, taux de prédation des défoliateurs, richesse en oiseaux	Variables LiDAR

Pour cela, nous utilisons les fonctions du paquet « mixOmics » sous **R**. La fonction pls() permet la modélisation puis nous pouvons dessiner le cercle de corrélation entre les variables et les individus sur les composantes principales. Ensuite, nous pouvons évaluer le pouvoir prédictif du modèle construit en effectuant une validation croisée. En raison du faible nombre d'observations

terrain dont nous disposons, nous effectuons cette validation en mode « leave one out » (i.e. nous apprenons sur n-1 parcelles et nous évaluons la déviance du modèle sur la n-ième et ceci est répété n fois). A partir de ce calcul, nous obtenons les valeurs moyennes d'erreur MSEP (Mean Square Error of Prediction) pour chaque composante ainsi que le Q² (un équivalent du R² mais en prédictif). Nous pouvons ainsi connaître le nombre de composantes utiles et évaluer la pertinence du modèle. A noter que nous considérons que le Q² doit être supérieur à la valeur ad hoc de 0.0975 pour valider le modèle. Les résultats obtenus seront à nouveau détaillés dans la section IV de cette partie.

Concernant les résultats, le modèle obtenu n'est pas significatif (voir figures 25 et 26 et scripts R ci-dessous). En effet, la corrélation entre les variables réponses sur les deux composantes principales est faible. De plus, le pouvoir prédictif du modèle est non significatif comme en atteste les valeurs du MSEP et du Q² calculées. Le MSEP est proche de 1 ce qui signifie que le modèle n'explique pas la variance de la variable réponse. Le Q² est aussi inférieur à la valeur ad hoc validant la prédiction. Dans cette étude, la modélisation PLS n'est donc pas satisfaisante. Les avantages de la méthode, à savoir la possible corrélation des variables utilisées, l'utilisation de plusieurs variables réponses et l'évaluation du pouvoir de prédiction du modèle, ne sont donc pas utiles ici.

```
varLid <- data.frame(dataMetricLidIBPMuesliUseful, autresVarLid)
varReponse <-
data.frame(as.numeric(dataBoisUseful$ic_global), dataChenilleUseful$pred_oiseau,
dataBoisUseful$ri_sp_ois, dataBoisIBPUseful$ibp)
pls.reg <- pls(varLid, varReponse, ncomp=4, mode="regression") #varLid, puis
responses
plotVar(pls.reg, comp=1:2, style = "lattice", overlap = FALSE)
```



```

tune.pls <- perf(pls.reg, validation = 'loo', criterion = 'all', progressBar =
FALSE)
# Leave one out : on apprend sur n-1 parcelles et on regarde sur la n ieme
parcelle de combien on s'est trompe

round(tune.pls$MSEP, 2)

##                1 comp 2 comp 3 comp 4 comp
## as.numeric.dataBoisUseful.ic_global.  0.97  1.00  1.19  1.21
## dataChenilleUseful.pred_oiseau      1.06  1.14  1.14  1.21
## dataBoisUseful.ri_sp_ois             0.95  0.97  1.05  1.11
## dataBoisIBPUseful.ibp                1.00  0.93  0.97  1.09

#MSEP : mean square error of prediction à comparer aux autres modèles et voir
Lequel est le meilleur
round(tune.pls$Q2, 2)

##                1 comp 2 comp 3 comp 4 comp
## as.numeric.dataBoisUseful.ic_global.  0.07 -0.04 -0.08 -0.07
## dataChenilleUseful.pred_oiseau      -0.02 -0.08 -0.01 -0.09
## dataBoisUseful.ri_sp_ois            0.09 -0.02 -0.10 -0.09
## dataBoisIBPUseful.ibp               0.03  0.08 -0.09 -0.16

# L'équivalent du R2, mais en prédictif
round(tune.pls$Q2.total, 2)

##          Q2.total
## 1 comp    0.04
## 2 comp   -0.02
## 3 comp   -0.07
## 4 comp   -0.10

# Le Q2 total (toutes les variables)

plot(tune.pls$Q2.total)
abline(h = 0.0975) # La valeur ad hoc au-dessus de laquelle on considère que le
Q2 est bon

```

Glossaire et sigles utiles

ACP	Analyse en Composantes Principales, méthode de statistiques multivariées
AIC	Akaike Information Criterion, mesure de la qualité d'un modèle statistique
CEFS	Comportement et Ecologie de la Faune Sauvage
Dynafor	DYNamiques et écologie des paysages AgriFORestiers
ENSAT	Ecole Nationale Supérieure Agronomique de Toulouse
GLM	Generalized Linear Model, modèle statistique généralisant la regression linéaire
GPS	Global Positioning System, système américain de positionnement par satellites
IBP	Indice de Biodiversité Potentielle, outil d'intégration de la biodiversité dans la gestion forestière
INRA	Institut National de Recherche Agronomique
LiDAR	Light Detection And Ranging, technique de mesure de distance à partir d'un faisceau laser
MEA	Millenium Ecosystem Assessment, programme mondial d'évaluation de l'impact humain sur les écosystèmes
MNH	Modèle Numérique de Hauteur, différence entre MNS et MNT
MNS	Modèle Numérique de Surface, représentation du sol et du sursol
MNT	Modèle Numérique de Terrain, représentation de la topographie d'une zone
MUESLI	MULTiscale mapping of Ecosystem Services by very high spatial resolution hyperspectraL and LiDAR remote sensing Imagery
NDVI	Normalized Difference Vegetation Index, indice normalisé caractérisant la couverture végétale
PLS	Partial Least Squares, méthode statistique de régression par les moindres carrés partiels
(LiDAR) ROC	LiDAR à Retour d'Onde Complète, type de LiDAR permettant d'enregistrer l'intégralité du signal rétrodiffusé
SIG	Système d'Information Géographique, système d'information pour données spatialisées
VCI	Vertical Complexity Indice, indice de diversité des différentes strates de la végétation

Table des matières

Introduction	8
1. Présentation du contexte	9
I. Organisme d'accueil	9
II. Projet interdisciplinaire MUESLI	10
1) Estimation d'indicateurs de biodiversité dans des paysages agriforestiers.....	10
2) Apport potentiel de la télédétection	10
III. Problématique et objectifs du stage	11
IV. Organisation du travail	11
2. LiDAR et écologie : état de l'art.....	12
I. Rappel du principe d'un levé LiDAR aéroporté	12
II. Usage des données LiDAR pour des questions écologiques.....	14
1) Apport du LiDAR en écologie et notamment foresterie.....	14
2) Apport du LiDAR pour l'évaluation des services écosystémiques	16
III. Variables LiDAR usuellement rencontrées dans la littérature permettant de décrire le nuage de points.....	18
1) Extraction de données du nuage de points	18
2) Caractérisation du nuage de points LiDAR à l'échelle de l'arbre	21
IV. Principaux logiciels libres de traitement des données LiDAR	22
3. Matériels	24
I. Site d'étude : les coteaux de Gascogne.....	24
II. Forêts échantillonnées et données écologiques associées.....	24
III. Données LiDAR (et hyperspectrales)	27
IV. Sortie terrain	29
4. Modélisation de variables écologiques à partir de variables LiDAR.....	30
I. Démarche d'extraction des variables LiDAR.....	30

II. Modélisation statistique	34
1) Exploration des données	34
2) Modélisation statistique par GLM.....	38
III. Prédiction et spatialisation des variables écologiques	40
IV. Résultats.....	41
V. Discussion	50
5. Extraction automatique des cimes et couronnes des arbres.....	56
I. Intérêts et méthodes existantes	56
II. Démarche adoptée et mise en œuvre.....	57
III. Résultats	59
IV. Discussion	65
Conclusion	67

Liste des tableaux

Tableau 1 Variables dérivées de l'analyse de détection des cimes tiré de Lesur (2011).	21
Tableau 2 Comparaison de logiciels permettant le traitement des données LiDAR.....	22
Tableau 3 Récapitulatif des variables écologiques à disposition et échelle des relevés terrains. ..	27
Tableau 4 Variables LiDAR peu corrélées à chaque échelle d'analyse.	37
Tableau 5 Modèles GLM calculés.	38
Tableau 6 Distribution des variables réponses.....	40
Tableau 7 Résultats de chaque modèle GLM obtenu : déviance, variables significatives, taux de significativité, sens des effets.	48
Tableau 8 Résultats de l'extraction par maxima locaux pour un filtrage gaussien 5*5 : distance euclidienne moyenne par rapport à la couche de validation, nombre de cimes valides parmi celles obtenues, erreur d'omission et de commission pour chaque rayon de recherche et par rapport à la distance maximale de validation d'une cime.	63
Tableau 9 Résultats de l'extraction par maxima locaux pour un filtrage gaussien 3*3 : distance euclidienne moyenne par rapport à la couche de validation, nombre de cimes valides parmi celles obtenues, erreur d'omission et de commission pour chaque rayon de recherche et par rapport à la distance maximale de validation d'une cime.	63
Tableau 10 Résultats de l'extraction par maxima locaux sans filtrage : distance euclidienne moyenne par rapport à la couche de validation, nombre de cimes valides parmi celles obtenues, erreur d'omission et de commission pour chaque rayon de recherche et par rapport à la distance maximale de validation d'une cime.	64
Tableau 11 Résultats de l'extraction par ligne de partage des eaux : distance euclidienne moyenne par rapport à la couche de validation, nombre de cimes valides parmi celles obtenues, erreur d'omission et de commission pour chaque filtrage.....	64
Tableau 12 Modèle PLS calculé.....	162

Liste des figures

Figure 1 Schéma d'organisation de l'UMR.	9
Figure 2 Axes de recherche de Dynafor, source : site internet du laboratoire.....	10
Figure 3 MNH raster d'un bois (à gauche) et nuage de points LiDAR colorisé suivant l'altitude (à droite).....	13
Figure 4 Classe végétation d'un nuage de points LiDAR, colorisé suivant l'altitude des points.	13
Figure 5 Schéma de l'approche par zone tiré de White et al. (2013) : calibration de modèles prédictifs à partir des relevés terrains des placettes et construction d'une grille dans laquelle les variables LiDAR obtenues par balayage laser aéroporté (BLA) vont être extraites pour chaque cellule de la taille des placettes afin de prédire la variable dendrométrique d'intérêt sur toute la zone d'étude.	16
Figure 6 Schéma du lien entre les écosystèmes et l'homme rempli par les services écosystémiques adapté d'UICN France (2012).	17
Figure 7 Schéma expliquant le calcul des percentiles tiré de David Herries (2014).....	19
Figure 8 Schéma de différenciation des différents retours LiDAR (proportions approximatives) tiré de l'université d'Humboldt.....	19
Figure 9 Illustration du calcul de la transparence de canopée adaptée du manuel d'utilisation de FUSION, source : http://forsys.cfr.washington.edu/fusion/FUSION_manual.pdf).....	20
Figure 10 Illustration du calcul de la couverture de canopée adaptée du manuel d'utilisation de FUSION.	20
Figure 11 Localisation du site d'étude.	24
Figure 12 Protocole des relevés du taux de prédation des chenilles.	27
Figure 13 Zone des levés LiDAR et situation des bois relevés.....	28
Figure 14 Caractéristiques des levés LiDAR issues du cahier des charges.	28
Figure 15 Schéma des étapes de la modélisation statistique, de l'extraction des données LiDAR à la spatialisation des variables écologiques.	30
Figure 16 Processus d'extraction des variables LiDAR.	31
Figure 17 Exemple d'un bois coupé par une ligne électrique et ayant dû être retracé.	32
Figure 18 Schéma des étapes d'exploration des données à réaliser avant une modélisation statistique.....	36
Figure 19 Boîte à moustache de la prédation des défoliateurs à l'échelle des bois.	41

Figure 20	Corrélogramme des variables peu corrélées à l'échelle des bois (croix noire : valeur non significative, taille des cercles proportionnelle aux coefficients de corrélation).	42
Figure 21	Boîtes à moustaches des différentes variables LiDAR, centrées et réduites, peu corrélées à l'échelle des bois.	44
Figure 22	Diagrammes de dispersion du taux de prédation des chenilles en fonction des variables LiDAR peu corrélées à l'échelle des plots.	46
Figure 23	Visualisation des données LiDAR pour deux plots présentant des valeurs extrêmes de VCI.	51
Figure 24	Photographie terrain et coupe verticale du nuage de points.	51
Figure 25	Prédiction de l'IBP à partir de variables LiDAR et de la variance du NDVI (correspondant au modèle créé n°2) sur 11 bois de la zone des levés.	55
Figure 26	Photographie aérienne de la zone d'étude, issue de Google maps (à gauche) et entité vectorielle dessinée à partir du MNH (à droite).	58
Figure 27	Couche vectorielle créée avec l'emplacement de chaque cime.	58
Figure 28	Résultat avec la méthode des maxima locaux (script Python) et un filtrage gaussien 3*3 rayon de recherche de 3.5m.	60
Figure 29	Résultat avec la méthode des maxima locaux (script Python) et un filtrage gaussien 3*3 rayon de recherche de 4m.	60
Figure 30	Résultat avec la méthode des maxima locaux (script Python) et un filtrage gaussien 3*3 rayon de recherche de 4.5m.	61
Figure 31	Résultat avec FUSION et un filtrage gaussien 3*3.	61
Figure 32	Résultat avec FUSION et un filtrage gaussien 5*5.	62
Figure 33	Résultat avec FUSION sans filtrage.	62
Figure 34	Diagramme de GANTT du stage.	73
Figure 35	Fiches de sensibilisation à l'IBP explicitant les facteurs entrant dans le calcul d'après Larrieu et Gonin.	75
Figure 36	Visualisation des données LiDAR pour deux plots présentant des valeurs extrêmes de la couverture de canopée.	76
Figure 37	Visualisation des données LiDAR pour deux plots présentant des valeurs extrêmes de l'inverse de la densité de canopée.	76
Figure 38	Visualisation des données LiDAR pour deux plots présentant des valeurs extrêmes de la proportion de 1ers retours de faible intensité.	76
Figure 39	Visualisation terrain/nuage de points LiDAR pour deux zones d'un profil vertical.	77
Figure 40	Script de commandes batch permettant de lancer l'extraction des variables LiDAR avec LAStools et FUSION à l'échelle de plots.	78

Figure 41 Script de commandes batch permettant de lancer l'extraction des variables LiDAR avec LAStools et FUSION par grille.....	78
Figure 42 Cercle des corrélations des variables explicatives sur les deux composantes principales.	164
Figure 43 Cercle des corrélations des variables réponses sur les deux composantes principales.	164

Liste des annexes

Annexe 1 : Diagramme de GANTT	73
Annexe 2 : Calcul de l'iBP	74
Annexe 3 : Visualisation des données LiDAR	76
Annexe 4 : Code d'extraction de variables LiDAR	78
Annexe 5 : Résultats de la modélisation statistique par GLM.....	79
Annexe 6 : Résultats de la segmentation	154
Annexe 7 : Modélisation statistique par PLS	162