

Apport du LiDAR aéroporté à la modélisation d'habitat de l'avifaune forestière wallonne : une approche comparative par Random Forest et sélection adaptative de variables.

Auteur : Pauly, Emma

Promoteur(s) : Lejeune, Philippe; 3917

Faculté : Gembloux Agro-Bio Tech (GxABT)

Diplôme : Master en bioingénieur : gestion des forêts et des espaces naturels, à finalité spécialisée

Année académique : 2024-2025

URI/URL : <http://hdl.handle.net/2268.2/24223>

Avertissement à l'attention des usagers :

Tous les documents placés en accès ouvert sur le site le site MatheO sont protégés par le droit d'auteur. Conformément aux principes énoncés par la "Budapest Open Access Initiative"(BOAI, 2002), l'utilisateur du site peut lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces documents, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale (ou prévue par la réglementation relative au droit d'auteur). Toute utilisation du document à des fins commerciales est strictement interdite.

Par ailleurs, l'utilisateur s'engage à respecter les droits moraux de l'auteur, principalement le droit à l'intégrité de l'oeuvre et le droit de paternité et ce dans toute utilisation que l'utilisateur entreprend. Ainsi, à titre d'exemple, lorsqu'il reproduira un document par extrait ou dans son intégralité, l'utilisateur citera de manière complète les sources telles que mentionnées ci-dessus. Toute utilisation non explicitement autorisée ci-avant (telle que par exemple, la modification du document ou son résumé) nécessite l'autorisation préalable et expresse des auteurs ou de leurs ayants droit.

**APPORT DU LIDAR AÉROPORTÉ À LA MODÉLISATION
D'HABITAT DE L'AVIFAUNE FORESTIÈRE WALLONNE :
*UNE APPROCHE COMPARATIVE PAR RANDOM FOREST ET
SÉLECTION ADAPTATIVE DE VARIABLES***

EMMA PAULY

TRAVAIL DE FIN D'ETUDES PRESENTE EN VUE DE L'OBTENTION DU DIPLOME DE
MASTER BIOINGENIEUR EN GESTION DES FORETS ET DES ESPACES NATURELS

ANNEE ACADEMIQUE 2024-2025

CO-PROMOTEURS : PHILIPPE LEJEUNE & JEAN-YVES PAQUET

© Toute reproduction du présent document, par quelque procédé que ce soit, ne peut être réalisée qu'avec l'autorisation de l'auteur et de l'autorité académique de Gembloux Agro-Bio Tech.

Le présent document n'engage que son auteur.

**APPORT DU LIDAR AÉROPORTÉ À LA MODÉLISATION
D'HABITAT DE L'AVIFAUNE FORESTIÈRE WALLONNE :
*UNE APPROCHE COMPARATIVE PAR RANDOM FOREST ET
SÉLECTION ADAPTATIVE DE VARIABLES***

EMMA PAULY

TRAVAIL DE FIN D'ETUDES PRESENTE EN VUE DE L'OBTENTION DU DIPLOME DE
MASTER BIOINGENIEUR EN GESTION DES FORETS ET DES ESPACES NATURELS

ANNEE ACADEMIQUE 2024-2025

CO-PROMOTEURS : PHILIPPE LEJEUNE & JEAN-YVES PAQUET

Résumé

La modélisation d'habitat aviaire s'appuie traditionnellement sur des variables de composition forestière et de configuration paysagère bidimensionnelles, qui peinent à caractériser la complexité structurale tridimensionnelle déterminante pour les espèces forestières. Cette étude quantifie l'apport des données LiDAR aéroporté dans la modélisation d'habitat de l'avifaune wallonne et identifie les variables structurales les plus informatives selon les profils écologiques.

Un pipeline adaptatif Random Forest a été développé pour 61 espèces aviaires utilisant les données SOCWAL (2019-2022), intégrant 28 métriques LiDAR, 18 variables de composition forestière et 32 variables paysagères. La sélection de variables par VSURF et la validation croisée spatiale ont été adaptées selon la prévalence des espèces (communes $\geq 15\%$, rares 5-15%, très rares $< 5\%$). L'apport LiDAR a été quantifié par comparaison des performances AUC-ROC entre modèles complets et modèles excluant les variables tridimensionnelles.

Les variables LiDAR améliorent significativement la modélisation pour 57% des espèces (gain médian d'AUC +0.016), avec une forte variabilité inter-spécifique. Huit espèces montrent un gain (≥ 0.05), incluant le Coucou gris (+0.0556) et la Fauvette des jardins (+0.0382). Contrairement aux hypothèses, les espèces spécialistes ne bénéficient pas systématiquement des variables LiDAR. L'écart-type des hauteurs de canopée (CHM_Height_StdDev_m) émerge comme la variable structurale la plus universellement informative (67% des espèces), tandis que les indices de complexité sophistiqués sont moins fréquemment sélectionnés.

L'apport du LiDAR à la modélisation d'habitat aviaire est espèce-spécifique plutôt qu'universellement bénéfique. Les variables structurales simples (hétérogénéité des hauteurs, couverture de canopée) priment sur les indices de complexité mathématique. Ces résultats orientent vers des stratégies de conservation différenciées et identifient l'hétérogénéité verticale comme déterminant structural prioritaire pour la gestion forestière wallonne.

Mots-clés : Avifaune forestière, LiDAR aéroporté, modélisation d'habitat, Random Forest, structure tridimensionnelle, validation croisée spatiale, VSURF, écologie aviaire, Wallonie

Abstract

Avian habitat modeling traditionally relies on two-dimensional forest composition and landscape configuration variables, which fail to characterize the three-dimensional structural complexity that is determinant for forest species. This study quantifies the contribution of airborne LiDAR data to avian habitat modeling in Wallonia and identifies the most informative structural variables according to ecological profiles.

An adaptive Random Forest pipeline was developed for 61 avian species using SOCWAL data (2019-2022), integrating 28 LiDAR metrics, 18 forest composition variables, and 32 landscape variables. Variable selection using VSURF and spatial cross-validation were adapted according to species prevalence (common $\geq 15\%$, rare 5-15%, very rare $< 5\%$). LiDAR contribution was quantified by comparing AUC-ROC performance between complete models and models excluding three-dimensional variables.

LiDAR variables significantly improve modeling for 57% of species (median AUC gain +0.016), with high inter-specific variability. Eight species show substantial gains (≥ 0.05), including Common Cuckoo (+0.0556) and Garden Warbler (+0.0382). Contrary to hypotheses, specialist species do not systematically benefit from LiDAR variables. Canopy height standard deviation (CHM_Height_StdDev_m) emerges as the most universally informative structural variable (67% of species), while sophisticated complexity indices are less frequently selected.

LiDAR contribution to avian habitat modeling is species-specific rather than universally beneficial. Simple structural variables (height heterogeneity, canopy cover) outperform mathematical complexity indices. These results guide toward differentiated conservation strategies and identify vertical heterogeneity as a priority structural determinant for Walloon forest management.

Keywords: Forest birds, airborne LiDAR, habitat modeling, Random Forest, three-dimensional structure, spatial cross-validation, VSURF, avian ecology, Wallonia

Remerciements

Je voudrais tout d'abord remercier mon promoteur Philippe Lejeune pour son temps, ses relectures attentives et ses précieux conseils qui ont guidé ce travail tout au long de son élaboration.

Ce travail n'aurait pas été possible sans l'aide du professeur Yves Brostaux dont l'expertise et les conseils ont été déterminants pour l'aboutissement de ce travail.

Un remerciement particulier à Nicolas Latte pour son aide précieuse concernant le traitement et l'analyse des données LiDAR, ainsi que pour ses conseils techniques qui ont permis de surmonter certains défis méthodologiques.

Ce travail n'aurait pas pu être réalisé sans Jean-Yves Paquet ainsi que tous les ornithologues passionnés qui ont collecté des données à travers toute la Wallonie depuis le début du projet SOCWAL.

Je tiens également à remercier les naturalistes et ornithologues avec qui j'ai pu échanger et discuter de ce travail, enrichissant ma compréhension de l'écologie aviaire.

Un grand merci à mes colocataires Roxane et Elisa (et Pipou) qui m'ont écoutée m'énerver quand le code ne fonctionnait pas et qui ont toujours été là pour décompresser et rigoler. Leur présence et leur bonne humeur ont rendu cette période incroyable.

Je voudrais également remercier Jimmy, technicien à la faculté, pour nos séances impromptues de lecture d'articles sur les oiseaux pendant ses scans racinaires. Ces moments d'échange scientifique informels ont contribué à nourrir ma réflexion.

Enfin, je souhaite rendre hommage à Nicole, qui m'a transmis sa passion pour la beauté et la simplicité de l'observation de la nature. Son sens aigu de l'observation et son amour pour les oiseaux du jardin m'ont influencée dans le choix de ce mémoire. Ses précieux conseils et enseignements m'ont orientée et guidée dans ce travail. Même si elle n'est plus là pour voir le travail fini, elle y a contribué et continue d'y contribuer encore maintenant dans la pérennité de ses enseignements.

Liste des Abréviations.....	5
1. INTRODUCTION.....	6
1.1 Contexte et enjeux.....	6
1.2 Problématique.....	7
1.3 Hypothèses.....	8
1.4 Objectifs.....	9
2. ÉTAT DE L'ART.....	10
2.1 Modélisation d'habitat aviaire.....	10
Évolution des approches méthodologiques.....	10
Variables traditionnelles et leurs limitations.....	11
Défis et Limitations de la Modélisation d'Habitat Aviaire.....	12
Variabilité Temporelle et Dynamique des Habitats.....	12
Transférabilité et Intégration de Nouvelles Technologies.....	13
Contraintes Computationnelles et Communication.....	13
2.2 Technologie LiDAR en écologie.....	14
Principes techniques du LiDAR aéroporté.....	14
Avantages et limitations.....	16
2.3 Structure d'habitat et préférences aviaires.....	17
Théorie écologique des relations structure-avifaune.....	17
Spécialisation écologique et guildes fonctionnelles.....	17
3. ZONE D'ÉTUDE ET DONNÉES.....	19
3.1 Présentation de la zone d'étude.....	19
3.2 Données aviaires.....	19
Programme SOCWAL et protocole de collecte.....	19
Méthodologie des points d'écoute.....	19
Données utilisées pour la modélisation.....	19
Avantages et limitations méthodologiques des données SOCWAL.....	20
Avantages méthodologiques.....	20
Limitations méthodologiques.....	20
Implications pour la modélisation.....	21
3.3 Données LiDAR.....	21
Campagne LiDAR 2021-2022 - Wallonie.....	21
3.4 Variables environnementales complémentaires.....	22
3.4.1 Variables de composition forestière.....	22
3.4.2 Variables d'occupation et d'utilisation du sol.....	24
3.5 Random Forest : principe, avantages et limitations.....	25
Principe algorithmique.....	25

Avantages algorithmiques.....	25
Avantages spécifiques en écologie aviaire.....	26
Limitations et biais potentiels.....	26
Biais méthodologiques critiques.....	27
Considérations pour l'interprétation.....	27
3.6 VSURF : Principe, Avantages et Limitations.....	28
Principe algorithmique.....	28
Avantages de la méthode.....	28
Limitations et biais.....	28
Considérations d'interprétation.....	29
4. MÉTHODOLOGIE.....	29
4.1 Design d'étude et stratégie d'analyse.....	29
4.1.1 Approche générale.....	31
4.1.2 Pipeline adaptatif selon la prévalence.....	31
4.2 Données environnementales et prétraitement.....	31
4.2.1 Calcul des variables.....	31
Variables de composition forestière (n = 18) :.....	33
Variables paysagères (n = 32) :.....	33
4.2.2 Stratégie de génération des pseudo-absences.....	35
4.3 Caractérisation de l'autocorrélation spatiale.....	35
4.3.1 Diagnostic de l'autocorrélation.....	35
4.3.2 Paramètres d'autocorrélation extraits.....	36
4.4 Validation croisée spatiale adaptative.....	36
4.4.1 Principe de l'approche adaptative.....	36
4.4.2 Méthode 1 : Block Cross-Validation standard (espèces communes $\geq 15\%$).....	37
4.4.3 Méthode 2 : Validation stratifiée spatiale (espèces rares 5-15%).....	38
4.4.4 Méthode 3 : Leave-One-Out spatial (espèces très rares $< 5\%$).....	39
4.4.5 Évaluation de la qualité des folds.....	39
4.5 Sélection de variables par VSURF.....	40
4.5.1 Principe de l'algorithme VSURF.....	40
4.5.2 Configuration adaptative selon la prévalence.....	40
4.5.3 Validation de la stabilité des sélections.....	40
4.6 Modélisation par Random Forest.....	41
4.6.1 Justification du choix algorithmique.....	41
4.6.2 Paramétrage et optimisation.....	41
4.6.3 Procédure de validation croisée intégrée.....	41

4.7 Métriques d'évaluation des performances.....	42
4.7.1 Métriques primaires.....	42
4.7.2 Métriques complémentaires.....	42
4.8 Analyse comparative de l'apport LiDAR.....	42
4.8.1 Design comparatif.....	42
4.8.2 Tests de significativité statistique.....	43
4.8.3 Classification de l'apport LiDAR.....	43
4.9 Analyse de sensibilité et validation.....	43
4.9.1 Tests de robustesse.....	43
4.9.2 Analyse des résidus.....	43
4.10 Environnement computationnel et reproductibilité.....	44
4.10.1 Spécifications techniques.....	44
4.11 Limitations méthodologiques.....	44
4.11.1 Limitations inhérentes au design.....	44
4.11.2 Limitations algorithmiques.....	45
4.11.3 Considérations écologiques.....	45
4.11.4 Limitations des tests statistiques.....	45
5. RÉSULTATS.....	46
5.1 Caractérisation des espèces étudiées.....	46
5.2 Performance globale du modèle et apport des données LiDAR.....	47
Apport quantitatif des variables LiDAR.....	48
Variabilité selon la spécialisation écologique.....	49
Variables LiDAR universellement importantes.....	49
Patterns de sélection par guildes écologiques.....	51
5.4 Analyse critique des limitations méthodologiques.....	52
Biais inhérents au design spatial.....	52
Variabilité des configurations VSURF.....	52
Espèces problématiques identifiées.....	53
5.4 Validation écologique des résultats : cohérence avec les connaissances biologiques.....	53
Pic noir : spécialiste du bois mort et des vieux peuplements.....	54
Sittelle torchepot : cavicole des vieux chênes.....	54
Alouette des champs : spécialiste des milieux ouverts.....	54
Troglodyte mignon : insectivore de sous-bois.....	55
Validation de la cohérence des patterns inter-espèces.....	55
5.5 Synthèse critique et implications écologiques.....	56
Validation nuancée des hypothèses de recherche.....	56

Implications pour la gestion forestière.....	57
Espèces à intérêt conservatoire prioritaire.....	57
6. DISCUSSION.....	58
6.1 Interprétation écologique des résultats.....	58
Remise en question du paradigme spécialiste/généraliste.....	59
Prédominance des variables "simples" : la robustesse écologique prime.....	59
6.2 Comparaison avec la littérature internationale.....	60
Positionnement par rapport aux études similaires.....	60
Variables LiDAR importantes : convergence et spécificités régionales.....	60
6.3 Limites méthodologiques et autocritique.....	61
Limitations structurelles des données d'entrée.....	61
Choix méthodologiques : compromis et alternatives.....	61
Perfectibilité du modèle développé.....	62
6.4 Implications pour la gestion forestière.....	62
Recommandations opérationnelles prioritaires.....	62
Contexte réglementaire européen et indicateurs de suivi.....	63
Stratégies différenciées selon les profils écologiques.....	63
6.5 Perspectives méthodologiques et enjeux de recherche future.....	63
Améliorations du pipeline développé.....	64
Innovations technologiques et méthodologiques.....	64
Questions scientifiques émergentes.....	65
6.6 Portée et limites de transférabilité.....	65
Validité géographique et temporelle.....	65
Implications pour la recherche européenne.....	65
7. Contribution personnelle.....	66
8. CONCLUSION générale.....	67
RÉFÉRENCES BIBLIOGRAPHIQUES.....	68
Sites Internet et Ressources Numériques.....	75
Outils d'Intelligence Artificielle.....	75
Logiciels et Environnements de Développement.....	76
ANNEXES.....	77

Liste des Abréviations

- ❖ **AUC** : Area Under the Curve (Aire sous la courbe ROC)
- ❖ **BlockCV** : Block Cross-Validation (Validation croisée par blocs spatiaux)
- ❖ **CHM** : Canopy Height Model (Modèle de hauteur de canopée)
- ❖ **CRR** : Canopy Relief Ratio (Ratio de relief de canopée)
- ❖ **CV** : Cross-Validation (Validation croisée)
- ❖ **HIQR** : Height Interquartile Range (Écart interquartile des hauteurs)
- ❖ **LAD** : Leaf Area Density (Densité foliaire)
- ❖ **LiDAR** : Light Detection And Ranging (Détection et télémétrie par ondes lumineuses)
- ❖ **MDA** : Mean Decrease Accuracy (Diminution moyenne de la précision)
- ❖ **mtry** : Nombre de variables échantillonnées aléatoirement à chaque division
- ❖ **ntree** : Number of trees (Nombre d'arbres)
- ❖ **QGIS** : Système d'information géographique quantique
- ❖ **RF** : Random Forest (Forêt aléatoire)
- ❖ **ROC** : Receiver Operating Characteristic (Caractéristique de fonctionnement du récepteur)
- ❖ **SHP** : Shapefile (Format de fichier de données spatiales vectorielles)
- ❖ **SIG** : Système d'Information Géographique
- ❖ **SOCWAL** : Surveillance de l'Avifaune Commune en Wallonie
- ❖ **SPW** : Service Public de Wallonie
- ❖ **TIFF** : Tagged Image File Format (Format de fichier image balisé)
- ❖ **VSURF** : Variable Selection Using Random Forests
- ❖ **WALOUS** : Utilisation du Sol en Wallonie (WALlonia Land cOver and land USe)

1. INTRODUCTION

1.1 Contexte et enjeux

La forêt vue du ciel : Et si nous changions de point de vue pour mieux comprendre les habitats aviaires ?

Chaque matin, dans les forêts wallonnes, des centaines d'espèces d'oiseaux orchestrent une symphonie invisible. Pourtant, derrière cette apparente harmonie se cache une réalité préoccupante : de nombreuses espèces forestières voient leurs populations décliner, victimes d'un habitat que nous pensions bien connaître mais que nous ne savons finalement que difficilement caractériser.

Traditionnellement, la modélisation d'habitat aviaire s'appuie sur des cartographies forestières bidimensionnelles qui nous renseignent sur *quoi* pousse dans nos forêts – hêtres, chênes, épicéas – mais qui restent muettes sur *comment* ces arbres structurent l'espace tridimensionnel que les oiseaux habitent réellement.

Aujourd'hui, une révolution technologique silencieuse survole nos forêts : le LiDAR aéroporté. Ces « yeux laser » percent la canopée et révèlent, point par point, presque centimètre par centimètre, la complexité architecturale tridimensionnelle que nos cartes traditionnelles ne soupçonnaient pas. Mais cette technologie révolutionnaire est-elle essentielle pour la compréhension des habitats et donc la conservation aviaire ?

Cette recherche interroge l'apport réel des données LiDAR dans la modélisation d'habitat de l'avifaune wallonne. Elle confronte méthodes traditionnelles et technologies LiDAR sur 61 espèces d'oiseaux, en questionnant nos approches actuelles et explorant les voies d'une gestion forestière véritablement adaptée aux exigences écologiques de nos communautés aviaires.

La forêt vue du ciel peut-elle améliorer notre compréhension des habitats aviaires ?

1.2 Problématique

La modélisation de l'habitat aviaire s'appuie traditionnellement sur des variables environnementales décrivant la composition forestière (essences dominantes, diversité spécifique) et la configuration paysagère (fragmentation, connectivité). Bien que ces approches aient permis des avancées significatives dans la compréhension des relations espèce-habitat (Wintle et al., 2005), elles présentent des limitations importantes pour caractériser finement les préférences écologiques des espèces forestières.

Les variables de composition forestière, dérivées des inventaires forestiers classiques, renseignent sur la nature des essences présentes mais peinent à capturer l'hétérogénéité structurelle tridimensionnelle qui détermine la disponibilité des niches écologiques (Wintle et al., 2005). Les métriques paysagères, calculées à partir de données d'occupation du sol bidimensionnelles, caractérisent la configuration spatiale des habitats (Ferrier et al., 2002), mais ne rendent pas compte de la complexité architecturale intra-forestière. Cette limitation est particulièrement critique pour les espèces aviaires, dont l'utilisation de l'espace forestier est fortement déterminée par la stratification verticale et l'hétérogénéité structurelle horizontale (MacArthur & MacArthur, 1961 ; Whelan, 2000).

Les données LiDAR aéroporté offrent un potentiel prometteur pour dépasser ces limitations en permettant une caractérisation fine de la structure tridimensionnelle de la végétation (Vierling et al., 2008). Ces données permettent de calculer des métriques inédites caractérisant la complexité structurelle (indice de rugosité, diversité de hauteur du feuillage : Parker & Russ, 2004), la distribution verticale de la biomasse (densité foliaire par strate), les patterns d'ouverture de la canopée (ratio de trouées) et l'hétérogénéité spatiale (coefficient de variation des hauteurs). Cependant, l'intégration de ces variables dans la modélisation d'habitat aviaire reste peu explorée, et leur valeur ajoutée par rapport aux approches traditionnelles demeure à quantifier (McClure et al., 2012 ; Shaw et al., 2024).

Ce travail aborde trois questions de recherche fondamentales :

Première question : Les variables LiDAR apportent-elles une information significative à la modélisation d'habitat des espèces aviaires ? Cette question vise à quantifier l'amélioration des performances prédictives apportée par l'intégration des métriques structurelles dérivées du LiDAR par rapport aux variables traditionnelles de composition et de paysage.

Deuxième question : L'importance des variables LiDAR varie-t-elle selon les espèces, et quels types d'espèces sont les plus concernés ? McClure et al. (2012) ont démontré que certaines espèces forestières hivernantes nécessitent des informations de microhabitat pour une modélisation précise, suggérant une variabilité inter-spécifique dans les exigences structurales. Cette question explore si cette variabilité suit des patterns prévisibles selon les traits écologiques des espèces (spécialisation, guildes alimentaires, préférences de strate).

Troisième question : Quelles variables LiDAR sont les plus importantes pour caractériser l'habitat aviaire, et existe-t-il des variables universellement informatives ou spécifiques à chaque espèce ? L'identification de métriques structurales clés permettrait d'optimiser les protocoles d'acquisition de données et de développer des indicateurs opérationnels pour la gestion forestière.

Ces questions s'inscrivent dans une démarche d'écologie quantitative visant à développer des outils de modélisation plus précis et écologiquement pertinents pour la conservation de l'avifaune forestière wallonne.

1.3 Hypothèses

Hypothèse 1 (H1) : Les variables LiDAR améliorent significativement la prédiction d'habitat aviaire

Cette hypothèse s'appuie sur le principe établi par Whelan (2000) selon lequel la structure verticale complexe du feuillage accroît la tendance des individus à se déplacer entre habitats pour s'alimenter et éviter les menaces. Les métriques LiDAR, en caractérisant finement cette complexité structurale tridimensionnelle, devraient capturer des dimensions de l'habitat non accessibles aux variables traditionnelles, résultant en une amélioration mesurable des performances prédictives des modèles.

Hypothèse 2 (H2) : L'importance des variables LiDAR varie selon la spécialisation écologique des espèces

Cette hypothèse découle des observations de Fahrig (2003) sur l'importance différentielle de la superficie et de la configuration des parcelles selon les groupes biotiques. Les espèces spécialistes, ayant des exigences écologiques plus strictes, devraient montrer une sensibilité accrue aux variables structurales fines caractérisées par le LiDAR, tandis que les espèces

généralistes pourraient être moins dépendantes de ces métriques détaillées. Cette variabilité devrait suivre un gradient prévisible selon les traits écologiques (guildes alimentaires, préférences de strate, taille corporelle).

Hypothèse 3 (H3) : Les variables de complexité structurelle sont prépondérantes parmi les métriques LiDAR

Basée sur les travaux de Shaw et al. (2024) démontrant que l'hétérogénéité structurale des forêts améliore la richesse des oiseaux et la diversité acoustique, cette hypothèse prédit que les métriques quantifiant la complexité architecturale (rugosité de la canopée, diversité de hauteur du feuillage, hétérogénéité spatiale) seront systématiquement plus informatives que les métriques simples de hauteur ou de couverture. Cette prédominance refléterait l'importance écologique de l'hétérogénéité structurale pour la diversification des niches aviaires.

1.4 Objectifs

Objectif général

L'objectif concret de ce travail est de présenter les résultats d'un modèle capable, pour chaque espèce étudiée, d'identifier les variables principales caractérisant son habitat.

Les résultats du modèle pourront être utilisés pour mieux comprendre l'habitat de certaines espèces d'oiseaux actuellement en déclin. Ils permettront ainsi d'orienter et de fournir des pistes concrètes pour la restauration et/ou le maintien des habitats de ces espèces.

Objectifs spécifiques

1. Développer un pipeline méthodologique robuste pour l'intégration des variables LiDAR dans la modélisation d'habitat aviaire. Ce pipeline comprendra des stratégies adaptatives de sélection de variables selon la prévalence des espèces, une gestion rigoureuse de l'autocorrélation spatiale par validation croisée spatiale, et des protocoles standardisés d'évaluation comparative des performances.

2. Comparer l'importance relative des variables par catégorie en quantifiant la contribution respective des variables LiDAR, de composition forestière et paysagères à la performance prédictive des modèles.

3. Identifier les variables LiDAR les plus informatives par espèce cible, en caractérisant leur stabilité de sélection et leur importance écologique. L'analyse privilégiera les métriques montrant une robustesse inter-échantillons. Cette analyse sera complétée par une comparaison des résultats avec les connaissances écologiques existantes sur les espèces, afin de vérifier leur cohérence.

4. Analyser la variabilité inter-espèces et identifier les patterns d'importance des variables selon les traits écologiques. Cette analyse recherchera des relations entre la dépendance aux variables LiDAR et les caractéristiques biologiques (spécialisation, guildes, préférences de strate) pour établir des règles prédictives généralisables.

2. ÉTAT DE L'ART

2.1 Modélisation d'habitat aviaire

La modélisation d'habitat constitue un outil central de l'écologie de la conservation, définie par Wintle et al. (2005) comme "une représentation numérique des préférences d'une espèce en matière d'habitat". Ces modèles permettent de faire des inférences sur les besoins écologiques des espèces et leur réponse probable aux changements environnementaux, ou de prédire l'abondance, la densité et la probabilité d'occupation d'un site en fonction de ses caractéristiques environnementales..

Évolution des approches méthodologiques

L'évolution de la modélisation d'habitat aviaire a suivi trois générations principales d'approches. Les premiers modèles, développés par le Fish and Wildlife Service des États-Unis, reposaient sur les Indices de Qualité de l'Habitat (IQH) construits par expertise (Van Horne & Wiens, 1991). Ces modèles HSI représentent une approche conceptuelle reliant chaque variable environnementale à la pertinence d'un site pour l'espèce, sur une échelle de 0 (habitat non convenable) à 1 (conditions optimales) (Burgman et al., 2005). Bien que très flexibles et largement appliqués en gestion de la conservation (Reading et al., 1996 ; Breininger et al., 1998), leur faiblesse réside dans leur dépendance entière à la crédibilité de l'expert, rendant impossible leur évaluation statistique (Wintle et al., 2005).

La seconde génération a vu l'émergence des modèles statistiques, particulièrement les modèles de régression logistique (GLM) et les modèles additifs généralisés (GAM), qui se sont révélés écologiquement réalistes (Ferrier & Watson, 1997 ; Zaniwski et al., 2002). Les études comparatives ont démontré que la performance de la régression logistique est généralement au moins aussi bonne, sinon meilleure, que celle d'autres méthodes comme l'analyse des correspondances canoniques, les réseaux neuronaux ou les arbres de classification (Ferrier & Watson, 1997 ; Elith, 2000).

La troisième génération intègre les méthodes d'ensemble et d'apprentissage automatique, avec l'émergence de Random Forest comme approche de référence. Cette méthode non-paramétrique construit une multitude d'arbres de décision indépendants, chacun entraîné sur un échantillon aléatoire des données d'apprentissage, et agrège leurs prédictions par vote majoritaire (classification) ou moyenne (régression). Random Forest présente des avantages considérables : robustesse au sur-ajustement, gestion des interactions complexes et non-linéaires, capacité à traiter de nombreux prédicteurs corrélés, et fourniture de mesures d'importance des variables (Breiman, 2001). (La partie 3.5 [RandomForest](#) de ce travail apportera un développement plus détaillé sur ce sujet.)

Variables traditionnelles et leurs limitations

Les approches conventionnelles reposent sur deux catégories principales de variables. Les variables de composition caractérisent les attributs locaux des habitats : âge des forêts, composition floristique, densité de végétation,... Des indices spécialisés ont été développés à partir de cartes d'âge et de composition floristique pour caractériser la qualité d'habitat (Wintle et al., 2005). Les variables de configuration quantifient le contexte spatial : la proportion de forêts anciennes dans un rayon défini ou les métriques de fragmentation (Ferrier et al., 2002). Ces mesures de voisinage traduisent le contexte environnemental local, particulièrement important lorsque le domaine vital d'une espèce excède la taille des cellules de modélisation (Wintle et al., 2005).

Cependant, ces variables traditionnelles présentent des limitations importantes pour caractériser la complexité structurelle tridimensionnelle des écosystèmes forestiers. McClure et al. (2012) ont ainsi démontré l'échec de l'hypothèse selon laquelle "le macrohabitat seul est suffisant pour modéliser l'occupation de l'habitat par les oiseaux hivernant", soulignant la nécessité d'informations de microhabitat pour certaines espèces forestières.

Défis et Limitations de la Modélisation d'Habitat Aviaire

La modélisation d'habitat aviaire fait face à de nombreux défis méthodologiques et conceptuels qui limitent sa portée opérationnelle. L'autocorrélation spatiale constitue un problème fondamental : les points proches dans l'espace tendent à avoir des valeurs similaires, violant l'hypothèse d'indépendance des observations (Dormann et al., 2007). Cette violation peut conduire à une sous-estimation des erreurs standard et à une surévaluation de la significativité statistique des relations espèce-environnement. Les méthodes de validation croisée spatiale, notamment l'approche BlockCV (Valavi et al., 2019), ont été développées pour gérer cette problématique en créant des blocs géographiquement contigus, mais leur application reste techniquement exigeante.

La rareté des espèces pose des défis particuliers : fort déséquilibre entre présences et absences, instabilité des modèles due à la variance élevée des estimateurs, et risque de folds de validation sans présences (Breiner et al., 2015). Des approches spécialisées comme les méthodes de validation leave-one-out spatial ont été développées pour surmonter ces limitations (Lomba et al., 2010). Le choix des variables représente également un défi critique, particulièrement en présence de nombreuses variables potentiellement corrélées. Les algorithmes de sélection automatisée comme VSURF (Variable Selection Using Random Forests) ont été développés pour identifier les sous-ensembles optimaux tout en évitant le sur-ajustement (Genuer et al., 2015).

Variabilité Temporelle et Dynamique des Habitats

La variabilité temporelle des habitats constitue un enjeu majeur souvent négligé. Les besoins écologiques des oiseaux varient fortement selon les saisons (Faaborg et al., 2010), les stades du cycle de vie (nidification, migration, hivernage) (Newton, 2008), ou les conditions climatiques interannuelles (Both et al., 2006). La sélection d'habitat saisonnière illustre cette complexité : les besoins énergétiques durant la reproduction privilégient les habitats riches en arthropodes, tandis que les stratégies hivernales favorisent les ressources en graines et les abris thermiques (Holmes & Schultz, 1981).

Les variations climatiques interannuelles, comme l'Oscillation Nord-Atlantique (NAO), influencent significativement la phénologie forestière européenne et la disponibilité des ressources (Stenseth et al., 2002). Les années à printemps précoce modifient la synchronisation entre les pics d'abondance d'arthropodes et les besoins nutritionnels des jeunes, affectant la

sélection d'habitat (Both et al., 2006). Cette variabilité climatique induit des changements spatiaux qui ne peuvent être captés par des modèles basés sur des données statiques, introduisant une autocorrélation spatio-temporelle complexe (Cressie & Wikle, 2015).

Les modèles temporellement figés risquent de surestimer la stabilité des habitats et de produire des résultats biaisés vers une vision déterministe de la distribution des espèces (Araújo & Peterson, 2012). L'intégration de données temporelles à haute résolution reste limitée par les contraintes logistiques des suivis à long terme et les coûts d'acquisition (Lindenmayer & Likens, 2010).

Transférabilité et Intégration de Nouvelles Technologies

La transférabilité des modèles (leur capacité à être appliqués à d'autres régions) constitue un verrou majeur (Yates et al., 2018). Un modèle calibré localement peut présenter des performances dégradées ailleurs, en raison de différences écologiques, de structure paysagère, ou d'interactions biotiques (Jiménez-Valverde et al., 2011). Cette non-stationnarité spatiale limite l'usage des modèles en conservation à grande échelle (Dormann et al., 2012).

L'intégration des données LiDAR et de télédétection offre un potentiel considérable pour décrire la structure tridimensionnelle des habitats (Davies & Asner, 2014), mais pose des défis techniques substantiels. Leur volume considérable (téraoctets pour des acquisitions régionales) et l'extraction de métriques écologiquement pertinentes nécessitent des compétences spécialisées (Lefsky et al., 2002). L'application d'algorithmes comme VSURF sur des centaines de métriques LiDAR risque de sélectionner des variables statistiquement significatives mais écologiquement fallacieuses (Strobl et al., 2008).

Contraintes Computationnelles et Communication

Les contraintes computationnelles représentent un facteur limitant croissant (Reichman et al., 2011). Les méthodes de validation croisée spatiale combinées aux algorithmes de sélection peuvent nécessiter des milliers d'ajustements de modèles, créant des goulots d'étranglement computationnels (Schratz et al., 2019). L'équilibre entre complexité du modèle, capacité prédictive, et interprétabilité écologique constitue un compromis difficile (Breiman, 2001).

La communication des résultats aux gestionnaires constitue un défi final : les modèles doivent être intelligibles et utilisables pour des actions concrètes de conservation (Cook et al., 2013).

Comment expliquer les subtilités de l'autocorrélation spatiale, les limitations de transférabilité temporelle, ou les incertitudes liées à la variabilité des habitats ? La visualisation des incertitudes et le développement d'interfaces utilisateur intuitives constituent des voies prometteuses pour faciliter l'appropriation par les gestionnaires de terrain.

2.2 Technologie LiDAR en écologie

Principes techniques du LiDAR aéroporté

Le LiDAR (Light Detection and Ranging) aéroporté constitue une technique de télédétection active qui repose sur l'émission d'impulsions laser depuis une plateforme aérienne pour modéliser la surface terrestre en trois dimensions (Wehr & Lohr, 1999). Cette technologie utilise le principe du temps de vol (Time of Flight), où le capteur mesure le délai entre l'émission de l'impulsion laser et la réception de son écho réfléchi par une surface.

Chaque impulsion laser peut produire plusieurs échos, permettant de distinguer les différentes couches d'un environnement (canopée, sol, objets intermédiaires). Cette caractéristique rend le LiDAR particulièrement efficace pour modéliser la structure verticale des écosystèmes et pour produire des modèles numériques de terrain (MNT) à travers la végétation dense (Hyypä et al., 2018). Les systèmes modernes atteignent une précision verticale de 5 à 15 cm et une précision horizontale de 10 à 30 cm, avec des densités de points allant de 10 à 50 points/m² pour des vols classiques.

L'intégration avec le GPS et les systèmes de navigation inertielle (INS) permet de déterminer avec précision la position et l'orientation du capteur LiDAR lors de la collecte de données, améliorant la précision des données spatiales obtenues. Les algorithmes de traitement des données sont essentiels pour interpréter les données LiDAR, notamment pour séparer les échos de la végétation de ceux du sol afin de cartographier avec précision les caractéristiques topographiques.

Métriques forestières dérivées

Les données LiDAR permettent de calculer une gamme étendue de métriques caractérisant la structure forestière tridimensionnelle, offrant une description fine des habitats forestiers inaccessible par les méthodes traditionnelles. Dans le cadre de ce travail, nous utilisons l'approche ABA (Area-Based Approach), qui subdivise l'espace forestier en cellules régulières

et calcule des métriques statistiques agrégées pour chaque unité spatiale. Cette méthode produit des métriques raster qui agrègent l'information tridimensionnelle des nuages de points LiDAR en grilles régulières, permettant une analyse spatiale standardisée et une intégration aisée avec d'autres données géographiques. L'approche ABA présente l'avantage de fournir une caractérisation robuste de la structure forestière à l'échelle du peuplement, et dans le cas de notre étude de la parcelle de 300m de rayon autour du point d'écoute.

Métriques de hauteur : Les percentiles de hauteur (P95, P99), la hauteur maximale, médiane et moyenne caractérisent la distribution verticale de la végétation. Ces métriques simples mais robustes constituent souvent la base des modèles prédictifs de biomasse forestière (Næsset, 2002) et se révèlent fortement corrélées à la diversité aviaire (Goetz et al., 2007). Les percentiles élevés (P95-P99) sont particulièrement efficaces pour estimer la hauteur dominante tout en étant moins sensibles aux valeurs aberrantes que la hauteur maximale (Magnussen & Boudewyn, 1998). La distribution des hauteurs reflète l'hétérogénéité structurale, déterminant majeur de la capacité d'accueil des communautés d'oiseaux forestiers (MacArthur & MacArthur, 1961).

Métriques de densité : La densité foliaire (LAD - Leaf Area Density) quantifie la distribution verticale de la biomasse foliaire et constitue un prédicteur clé de la productivité primaire et de la disponibilité en ressources alimentaires (Lefsky et al., 1999). Le ratio de pénétration LiDAR mesure la proportion d'impulsions atteignant le sol, indicateur robuste de la densité du couvert et de l'accessibilité du sous-bois pour les espèces terrestres (Korhonen et al., 2011). Les densités de retours par strate (sous-bois, canopée intermédiaire, canopée supérieure) caractérisent la stratification verticale, permettant d'identifier les niches écologiques spécifiques à chaque guildes d'oiseaux (Vierling et al., 2008). Ces métriques se révèlent particulièrement pertinentes pour distinguer les préférences des espèces canopéennes, des espèces de sous-bois et des espèces généralistes (Bergen et al., 2009).

Métriques de complexité : L'indice de rugosité (Rumple Index) quantifie la complexité tridimensionnelle de la surface de la canopée par le rapport entre surface 3D réelle et surface 2D projetée, reflétant l'hétérogénéité micro-topographique cruciale pour les espèces spécialistes (Parker & Russ, 2004). L'indice de diversité de hauteur du feuillage (Foliage Height Diversity - Shannon) adapte l'indice de Shannon classique à la distribution verticale de la végétation (Willson, 1974), s'appuyant sur les travaux pionniers de MacArthur & MacArthur (1961) qui ont établi la relation fondamentale entre complexité structurale verticale et

diversité aviaire. L'entropie verticale et les coefficients de variation des hauteurs complètent cette caractérisation en quantifiant l'hétérogénéité structurale à différentes échelles spatiales (Listopad et al., 2015).

Métriques de couverture et d'ouverture : Le ratio de trouées (Gap Fraction Ratio) quantifie la proportion d'ouvertures dans la canopée et détermine la pénétration lumineuse, facteur critique pour la régénération forestière et la structuration des communautés de sous-bois (Morsdorf et al., 2006). L'ouverture de canopée (Canopy Openness) mesure la perméabilité lumineuse et influence directement la productivité du sous-étage et la disponibilité en arthropodes (Halme & Niemelä, 1993). Ces métriques sont cruciales pour caractériser les micro-habitats de lisière et les conditions lumineuses au sol, déterminants pour les espèces héliophiles et les communautés d'insectes associées (Davies & Asner, 2014). La fraction de couverture par strates permet d'identifier les structures forestières favorables aux différentes guildes trophiques et de nidification (Hinsley et al., 2002).

Avantages et limitations

Le LiDAR présente des avantages considérables : caractérisation objective et quantitative de la structure 3D, reproductibilité des mesures, capacité de pénétration du couvert végétal, et fourniture de métriques impossibles à obtenir par d'autres moyens. Les données permettent une analyse multi-échelle, du microhabitat local aux patterns paysagers.

Toutefois, cette technologie présente des limitations qu'il convient de maîtriser pour éviter tout biais. Le LiDAR impose plusieurs contraintes : coût élevé d'acquisition, forte sensibilité aux conditions météorologiques, et complexité des algorithmes de traitement. De plus, des biais potentiels liés aux conditions d'acquisition peuvent influencer la validité des résultats (saison, densité de points, angle de vue). La problématique temporelle est particulièrement critique : les acquisitions hivernales favorisent la pénétration laser chez les feuillus mais peuvent sous-estimer la complexité structurale en période de végétation active. Ces facteurs de variabilité, loin d'être anecdotiques, peuvent avoir un impact considérable sur les résultats des études : ce sont des paramètres essentiels à prendre en compte, surtout lorsqu'on travaille dans un contexte temporel et qu'on vise une reproductibilité des études ainsi qu'une comparabilité des résultats dans le temps et l'espace.

2.3 Structure d'habitat et préférences aviaires

Théorie écologique des relations structure-avifaune

La relation entre structure forestière et communautés aviaires repose sur des fondements théoriques solides établis par MacArthur & MacArthur (1961) avec leur hypothèse de diversité de hauteur du feuillage. Cette théorie postule que la diversité aviaire est positivement corrélée à la complexité structurelle verticale de la végétation, chaque strate offrant des niches écologiques distinctes. Whelan (2000) a étendu cette théorie en démontrant que la structure verticale complexe du feuillage accroît la tendance des individus à se déplacer entre habitats pour s'alimenter et éviter les menaces.

Spécialisation écologique et guildes fonctionnelles

L'analyse des préférences d'habitat révèle des patterns différentiels selon les guildes écologiques. La structure de la végétation est plus importante que la structure des patches dans le maintien de la diversité des espèces, confirmant l'hypothèse d'hétérogénéité d'habitat (Lin et al., 2020 ; Tews et al., 2004).

Généralistes versus spécialistes : Les espèces généralistes tolèrent une large gamme d'habitats forestiers, tandis que les spécialistes présentent des exigences écologiques strictes liées à des caractéristiques structurelles spécifiques (Clavel et al., 2011). Cette dichotomie influence la sensibilité des modèles d'habitat aux variables LiDAR : les modèles d'espèces spécialistes montrent une dépendance accrue aux métriques structurelles fines (Zellweger et al., 2013), tandis que ceux des généralistes répondent davantage aux variables paysagères. Les espèces spécialistes sont particulièrement susceptibles d'extinction face aux changements structurels (Devictor et al., 2008).

Préférences de strate : Les communautés aviaires s'organisent selon une stratification verticale marquée (Holmes & Recher, 1986). Chez les insectivores de canopée exploitant les arthropodes du feuillage, les modèles d'habitat révèlent des corrélations positives avec des métriques de hauteur et de densité foliaire (Simonson et al., 2014). La diversité des essences crée une hétérogénéité des ressources alimentaires déterminante : les chênes attirent une entomofaune riche, tandis que les conifères fournissent des graines appréciées des granivores (Holmes & Robinson, 1981 ; Stański et al., 2021). Les insectivores de sous-bois nécessitent des

habitats à forte densité végétative dans les strates inférieures, à la fois pour la recherche alimentaire et pour leur protection. Cela se traduit par une corrélation positive avec les métriques LiDAR de densité des retours dans ces strates. Enfin, l'écorce rugueuse des arbres matures héberge une plus grande diversité d'arthropodes, accessibles aux espèces grimpeuses (Lammertink, 2007).

Mécanismes écologiques sous-jacents : La complexité structurelle augmente la diversité des microhabitats et des communautés d'arthropodes proies (Gumede et al., 2022). Les arbres de gros diamètre agissent comme des structures clés du paysage, fournissant des sites de nidification diversifiés selon les guildes : cavités naturelles pour les cavernicoles, houppiers denses ou ouverts selon les préférences architecturales (Cockle et al., 2015). La stratification verticale permet des stratégies d'évitement différentiel des prédateurs, avec un effet de prédation aviaire deux fois plus élevé dans les sous-étages denses (Gunnarsson et al., 2011).

Réponses aux métriques LiDAR : Les recherches récentes confirment que l'hétérogénéité structurelle des forêts, quantifiable par LiDAR, améliore la richesse aviaire et la diversité acoustique (Shaw et al., 2024). Les insectivores de canopée répondent positivement à la hauteur moyenne du couvert et à la complexité verticale, tandis que les insectivores de sous-bois préfèrent les zones à forte couverture arbustive détectable par les métriques de densité LiDAR (Gumede et al., 2022).

Implications pour la conservation : Cette compréhension théorique des relations structure-fonction justifie l'hypothèse selon laquelle les variables LiDAR, en caractérisant précisément la complexité structurelle tridimensionnelle, devraient améliorer significativement la modélisation d'habitat aviaire, avec une variabilité prévisible selon la spécialisation écologique des espèces. Les espèces spécialistes, dépendantes de configurations structurelles spécifiques (cavités, rugosité d'écorce, stratification verticale), devraient montrer une sensibilité accrue aux variables LiDAR de complexité structurelle (indice de rugosité, diversité de hauteur du feuillage, entropie verticale, coefficient de variation des hauteurs) et aux métriques de densité par strate (densité foliaire, densité de retours par strate verticale), tandis que les généralistes répondraient davantage aux variables de composition et de paysage. Cette approche différentielle ouvre des perspectives prometteuses pour une gestion forestière adaptée aux exigences écologiques spécifiques de chaque guildes aviaire.

3. ZONE D'ÉTUDE ET DONNÉES

3.1 Présentation de la zone d'étude

La Wallonie présente une mosaïque paysagère combinant 530 000 hectares de forêts (31% du territoire) (Alderweireld et al., 2015), allant des hêtraies pures aux plantations d'épicéas, des prairies extensives aux cultures intensives et des zones rurales aux centres urbains. Cette configuration paysagère fait de la Wallonie un laboratoire naturel optimal pour quantifier rigoureusement l'apport des variables de structure tridimensionnelle dérivées du LiDAR dans la modélisation prédictive des communautés aviaires.

3.2 Données aviaires

Programme SOCWAL et protocole de collecte

Les données ornithologiques proviennent du programme SOCWAL (Suivi des Oiseaux Communs en Wallonie), initié en 1990 et coordonné par Aves/Natagora. Ce programme contribue au réseau européen PECBMS (*Pan European Common Birds Monitoring Scheme*) et constitue l'un des rares dispositifs de suivi standardisé à long terme disponibles pour la recherche (Derouaux & Paquet, 2018).

Méthodologie des points d'écoute

Le protocole repose sur des points d'écoute standardisés : chaque observateur effectue deux passages annuels (20 mars-1er mai et 10 mai-30 juin) sur une chaîne de 15 points géolocalisés. Tous les oiseaux vus ou entendus sont recensés pendant 5 minutes par point, dans des conditions météorologiques standardisées.

Données utilisées pour la modélisation

Depuis 1990, 311 chaînes ont été suivies (4 438 points d'écoute uniques) par plus de 230 observateurs formés. Pour cette étude, seules les données 2019-2022 ont été utilisées pour correspondre aux acquisitions LiDAR disponibles à cette période. (Derouaux, 2021).

Les données de présence/pseudo-absence des espèces par point d'écoute constituent la variable réponse pour la modélisation des relations habitat-espèces.

Avantages et limitations méthodologiques des données SOCWAL

Avantages méthodologiques

Le programme SOCWAL présente plusieurs atouts méthodologiques décisifs pour cette étude. La standardisation rigoureuse du protocole (points d'écoute de 5 minutes, deux passages annuels standardisés, formation des observateurs) garantit la reproductibilité et la comparabilité des observations sur l'ensemble du territoire wallon. Cette standardisation, couplée à une continuité temporelle exceptionnelle de plus de 30 ans, constitue un avantage rare en écologie aviaire, permettant de distinguer les variations naturelles des tendances à long terme.

La couverture spatiale extensive (4 438 points d'écoute uniques répartis sur 311 chaînes) offre une bonne représentativité géographique pour les analyses à l'échelle régionale, tandis que l'intégration au réseau européen PECBMS assure la compatibilité méthodologique avec les standards internationaux de surveillance aviaire. Cette dimension internationale renforce la généralisabilité potentielle des résultats aux écosystèmes forestiers tempérés européens.

Limitations méthodologiques

Plusieurs limitations intrinsèques doivent être considérées dans l'interprétation des résultats. La structure spatiale non-aléatoire constitue une limitation majeure : les données sont collectées sous forme de chaînes de 15 points d'écoute consécutifs, générant une distribution spatiale agrégée plutôt qu'aléatoire. Cette organisation en clusters spatiaux exacerbe l'autocorrélation spatiale et peut conduire à une sous-estimation de la variabilité environnementale effective, nécessitant des approches de validation croisée spatiale particulièrement rigoureuses.

La détectabilité différentielle constitue un second biais majeur : la méthodologie d'observation diurne par points d'écoute de 5 minutes privilégie intrinsèquement certaines guildes aviaires au détriment d'autres. Les rapaces nocturnes et les espèces discrètes (passereaux cryptiques, espèces peu vocalisatrices) sont systématiquement sous-représentés dans les données collectées comparativement aux passereaux forestiers vocalisant activement durant les créneaux d'observation matinaux. Cette sélectivité méthodologique limite de facto l'étude à un sous-ensemble taxonomique de l'avifaune wallonne, excluant certaines espèces. Néanmoins, le volume considérable de données collectées compense cette limitation en fournissant une base

de données robuste pour l'analyse des espèces effectivement détectables par cette méthodologie.

La dépendance à l'accessibilité des sites induit potentiellement des biais spatiaux, avec une possible sous-représentation des milieux difficiles d'accès (zones humides, forêts anciennes isolées, falaises). Les données binaires présence/pseudo-absence ne renseignent pas sur l'abondance locale. Enfin, la restriction temporelle à la période reproductive (mars-juin) exclut les phénomènes hivernaux et migratoires, potentiellement critiques pour certaines espèces. Cette limitation temporelle peut masquer des relations habitat-espèce importantes en dehors de la saison de nidification.

Implications pour la modélisation

Ces limitations justifient l'adoption d'approches méthodologiques robustes : validation croisée spatiale pour gérer l'autocorrélation induite par la structure en chaînes, sélection adaptative de variables selon la prévalence des espèces, et interprétation prudente des modèles pour les espèces à détectabilité problématique. La structure agrégée des données impose particulièrement l'utilisation de techniques de validation spatiale avancées pour éviter l'optimisme artificiel des estimations de performance. Malgré ces contraintes, la qualité exceptionnelle du jeu de données SOCWAL compense largement ces limitations, en faisant une référence pour la recherche en écologie aviaire Wallonne.

3.3 Données LiDAR

Campagne LiDAR 2021-2022 - Wallonie

Les acquisitions LiDAR se sont déroulées de l'Ouest vers l'Est entre 2021 et 2022, durant les périodes hivernales (1er janvier - 31 mars) pour optimiser la pénétration laser à travers la végétation caduque.

Caractéristiques principales :

Densité : 6,8 points/m² (amélioration d'un facteur 4,5 par rapport à la campagne 2013-2014)

l'amélioration d'un facteur 4,5 en densité de points rend les données des deux campagnes méthodologiquement incompatibles pour les analyses comparatives temporelles. (Cf. annexe A)

Précision altimétrique : 0,12 m

Précision planimétrique : 1 m (RMSE)

Système de coordonnées : ETRS89 / Lambert Belge 2008

Les spécifications techniques détaillées de cette campagne de vol sont consultables sur Wallonemap (paramètres de vol, caractéristiques des capteurs, format de livraison, schéma de classification des points).

L'ensemble des métriques LiDAR a été calculé sur des parcelles circulaires de 300 m de rayon centrées sur chaque point d'écoute SOCWAL. Cette échelle spatiale correspond approximativement à l'aire d'influence acoustique des espèces aviaires lors des recensements de 5 minutes, tout en intégrant l'hétérogénéité structurelle du paysage forestier environnant et l'espacement limité entre les points d'écoute.

Les variables calculées sont mentionnées dans la section 4.2 de la méthodologie et décrites en annexe B,C,D.

3.4 Variables environnementales complémentaires

3.4.1 Variables de composition forestière

Les variables de composition forestière ont été calculées à partir de la cartographie des essences forestières développée par Bolyn et al. (2022)

Cette cartographie haute résolution (10 m × 10 m) résulte d'une approche innovante d'intégration multi-sources combinant séries temporelles Sentinel-2, données LiDAR aéroportées et 2 847 placettes d'inventaire forestier géoréférencées. L'algorithme Random Forest optimisé pour la classification d'essences forestières a produit une précision globale de 85,3% (coefficient Kappa : 0.82).

La nomenclature adoptée distingue 10 classes d'essences adaptées à la diversité forestière wallonne et aux exigences de la modélisation d'habitat aviaire :

1. **Chênes** (*Quercus* spp.) : incluant chêne pédonculé (*Q. robur*) et sessile (*Q. petraea*)
2. **Hêtres** (*Fagus sylvatica*) : essence dominante des forêts de feuillus wallonnes
3. **Épicéas** (*Picea abies*) : principal résineux de production en Wallonie
4. **Pins** (*Pinus* spp.) : regroupant pin sylvestre (*P. sylvestris*) et pin noir (*P. nigra*)
5. **Douglas** (*Pseudotsuga menziesii*) : essence introduite en expansion
6. **Mélèzes** (*Larix* spp.) : principalement mélèze d'Europe (*L. decidua*)
7. **Bouleaux** (*Betula* spp.) : essences pionnières et de lisière
8. **Peupliers** (*Populus* spp.) : incluant les peupliers naturels et cultivars
9. **Autres essences ligneuses** : regroupant les essences minoritaires (érables, frênes, aulnes, etc.)
10. **Végétation non-ligneuse** : zones de régénération, clairières, landes

L'ensemble des métriques de composition forestière a été calculé sur des **parcelles circulaires de 300 m de rayon** centrées sur chaque point d'écoute SOCWAL, pour assurer la cohérence spatiale avec les métriques LiDAR. Cette approche permet de quantifier la composition en essences, les indices de diversité spécifique (Shannon, Simpson) et les variables de structure forestière à l'échelle d'analyse appropriée pour la modélisation d'habitat aviaire.

Limitations principales

Plusieurs limitations méthodologiques doivent être considérées : la classification se limite aux essences principales wallonnes, excluant les espèces forestières minoritaires mais écologiquement importantes (érables, frênes, aulnes). L'approche pixellaire identifie uniquement l'essence dominante sans distinguer les mélanges complexes infra-pixellaires. La cartographie représente un état statique de 2020 sans intégrer les perturbations forestières récentes. Enfin, un biais d'échantillonnage potentiel découle de la répartition des placettes de référence, privilégiant les forêts de production au détriment des forêts naturelles ou des réserves, pouvant affecter la représentativité de la classification pour ces milieux spécifiques.

Pertinence écologique

Malgré les limitations évoquées, cette cartographie haute résolution constitue un progrès méthodologique majeur pour la modélisation d'habitat aviaire comparée aux approches traditionnelles. La précision taxonomique offerte par la distinction de 10 classes d'essences permet de tester des hypothèses écologiques fines sur les préférences spécifiques des espèces aviaires, impossible avec les classifications grossières d'occupation du sol habituellement

disponibles. Ceci apporte des informations précieuses pour identifier les associations spécifiques espèce-essence, particulièrement cruciales pour comprendre les mécanismes de sélection d'habitat des spécialistes forestiers.

Cette cartographie complète de manière optimale les métriques structurales dérivées du LiDAR.

Accessibilité des données : La cartographie est accessible via la plateforme ForestImator (<https://forestimator.gembloux.ulg.ac.be/>) développée par l'Université de Liège.

Les spécifications techniques détaillées, les performances par classe d'essence et l'analyse complète des limitations sont présentées dans l'article de Bolyn et al. (2022)

3.4.2 Variables d'occupation et d'utilisation du sol

Les variables paysagères ont été calculées à partir de la cartographie de l'utilisation du sol de Wallonie (WAL_UTS__2018) produite dans le cadre du projet WALOUS par un consortium SPW-UCLouvain/ULB/ISSEP (Service public de Wallonie et al., 2020).

Cette cartographie vectorielle haute résolution (1 m) couvre l'ensemble du territoire wallon pour l'année 2018 et caractérise l'usage des sols par parcelle cadastrale selon la définition INSPIRE. Elle résulte du croisement automatisé d'une vingtaine de géodonnées de référence garantissant une caractérisation exhaustive de l'utilisation du sol wallon.

L'ensemble des variables paysagères a été calculé sur des parcelles circulaires de 300 m de rayon centrées sur chaque point d'écoute SOCWAL, assurant la cohérence spatiale avec les métriques LiDAR et de composition forestière.

Cette approche permet de quantifier les métriques de diversité d'habitats et pressions anthropiques.

Pertinence écologique

Ces variables paysagères permettent de caractériser l'habitat aviaire à plusieurs échelles spatiales complémentaires de la structure fine (variables LiDAR) et de la composition forestière locale. Elles fournissent des informations cruciales sur la disponibilité et la connectivité des habitats préférentiels selon les exigences spécifiques de chaque espèce, ainsi que sur l'intensité des pressions anthropiques susceptibles d'influencer les patterns de

distribution aviaire (urbanisation, agriculture, sylviculture, etc.). La configuration spatiale des éléments paysagers, ajoutée aux métriques LiDAR et à la composition forestière, offre une vision intégrée des déterminants environnementaux de la distribution aviaire, de la microstructure forestière aux patterns paysagers régionaux.

Accessibilité des données : Géoportail de la Wallonie - <https://geodata.wallonie.be/>

Les spécifications techniques détaillées, la méthodologie de production complète et les considérations de qualité sont présentées sur le Géoportail de la Wallonie.

3.5 Random Forest : principe, avantages et limitations

Principe algorithmique

Random Forest est un algorithme d'ensemble non-paramétrique développé par Breiman (2001) qui combine les prédictions de multiples arbres de décision indépendants. L'algorithme repose sur deux sources de randomisation : le bootstrap aggregating (bagging) et la sélection aléatoire de variables. Pour chaque arbre, un échantillon bootstrap est tiré avec remise à partir des données d'entraînement, puis à chaque nœud de division, seul un sous-ensemble aléatoire de variables (typiquement \sqrt{p} pour la classification, où p est le nombre total de variables) est considéré pour déterminer la division optimale. La prédiction finale résulte du vote majoritaire des arbres individuels pour la classification ou de leur moyenne arithmétique pour la régression.

Cette double randomisation confère à Random Forest ses propriétés statistiques remarquables : chaque arbre présente un biais élevé mais une variance faible, et leur agrégation réduit drastiquement la variance globale sans augmenter significativement le biais, conformément au théorème de décomposition biais-variance (Hastie et al., 2009). L'algorithme génère également des estimations d'erreur out-of-bag (OOB) en testant chaque observation sur les arbres n'ayant pas utilisé cette observation dans leur entraînement, fournissant une estimation non-biaisée de l'erreur de généralisation sans validation croisée explicite.

Avantages algorithmiques

Random Forest présente des avantages considérables pour la modélisation écologique. Sa robustesse au sur-ajustement constitue un atout majeur : contrairement aux arbres individuels,

Random Forest reste stable même avec un grand nombre d'arbres, permettant d'augmenter le nombre d'estimateurs sans risque de dégradation des performances (Oshiro et al., 2012).

L'algorithme gère naturellement les interactions complexes et les relations non-linéaires sans spécification préalable, particulièrement pertinent pour les systèmes écologiques où les relations espèce-environnement sont souvent non-monotones.

La capacité à traiter de nombreux prédicteurs corrélés représente un avantage critique en écologie, où la multicolinéarité est fréquente. Random Forest n'exige aucune transformation préalable des variables et reste performant même lorsque le nombre de variables excède le nombre d'observations (Strobl et al., 2008). L'algorithme fournit des mesures d'importance des variables particulièrement utiles pour l'interprétation écologique : l'importance par Mean Decrease Accuracy (MDA) quantifie la dégradation de performance lorsqu'une variable est permutée aléatoirement, tandis que l'importance par Mean Decrease Gini quantifie la contribution moyenne de chaque variable à la pureté des nœuds.

Avantages spécifiques en écologie aviaire

Pour la modélisation d'habitat aviaire, Random Forest présente des avantages spécifiques. L'algorithme gère efficacement les données déséquilibrées fréquentes pour les espèces rares, via des mécanismes de pondération des classes et d'échantillonnage stratifié (Chen et al., 2004). Sa robustesse aux valeurs aberrantes est particulièrement pertinente pour les données environnementales qui peuvent présenter des mesures extrêmes liées à des artefacts de télédétection ou des conditions locales exceptionnelles.

La parallélisation native de Random Forest permet de traiter efficacement les grands jeux de données spatiaux typiques des études paysagères, tandis que sa capacité à quantifier l'incertitude des prédictions via la variance des votes individuels fournit des informations intéressantes pour la gestion conservatoire.

Limitations et biais potentiels

Malgré ses avantages, Random Forest présente des limitations importantes à considérer. L'algorithme montre un biais vers les variables catégorielles à nombreuses modalités et vers les variables continues à forte variabilité, pouvant artificiellement favoriser certains types de prédicteurs dans les mesures d'importance (Strobl et al., 2007). Ce biais est particulièrement

problématique lors de la comparaison entre variables LiDAR (continues à haute variabilité) et variables de composition (souvent catégorielles à faible variabilité).

L'interprétabilité limitée constitue une faiblesse majeure : contrairement aux modèles linéaires, Random Forest ne fournit pas de coefficients directement interprétables, rendant difficile la compréhension des mécanismes écologiques sous-jacents. L'algorithme peut également présenter un sur-ajustement aux données d'entraînement lorsque les arbres individuels sont très profonds, particulièrement avec des échantillons de petite taille typiques des espèces rares.

Biais méthodologiques critiques

Plusieurs biais méthodologiques peuvent affecter la validité des résultats. Le biais d'échantillonnage spatial est critique : si les données présentent une autocorrélation spatiale non prise en compte, Random Forest peut apprendre des patterns spatiaux artificiels plutôt que des relations écologiques authentiques (Roberts et al., 2017). Ce biais nécessite impérativement l'utilisation de techniques de validation croisée spatiale.

Le biais de sélection des variables peut également être problématique : Random Forest favorise les variables fortement corrélées aux vraies variables causales, même si elles ne sont pas causales elles-mêmes. Dans le contexte de cette étude, cela pourrait conduire à surestimer l'importance de variables LiDAR corrélées aux vraies variables structurelles déterminantes.

Considérations pour l'interprétation

L'interprétation des mesures d'importance doit être nuancée. La MDA (aussi appelée importance par permutation) peut être artificiellement élevée pour des variables bruitées dans des échantillons de petite taille, tandis que des variables écologiquement importantes mais corrélées peuvent présenter une importance réduite par effet de masquage mutuel. La stabilité des mesures d'importance doit donc être évaluée via des procédures de bootstrap ou de validation croisée répétée.

Enfin, Random Forest peut présenter des performances variables selon les espèces : les espèces à relations habitat simples peuvent être mieux modélisées par des approches paramétriques, tandis que Random Forest excelle pour les espèces à écologie complexe. Cette variabilité

inter-spécifique justifie une approche adaptative selon les caractéristiques écologiques des espèces cibles.

3.6 VSURF : Principe, Avantages et Limitations

Principe algorithmique

Variable Selection Using Random Forests (VSURF) est un algorithme de sélection de variables développé par Genuer et al. (2015) qui suit une approche séquentielle en trois étapes exploitant l'importance de permutation Random Forest :

- 1) élimination des variables non-informatives par comparaison avec des variables "bruit"
- 2) sélection pour l'interprétation via minimisation de l'erreur Out-of-Bag,
- 3) identification du sous-ensemble minimal pour la prédiction par élimination backward.

Avantages de la méthode

VSURF présente plusieurs atouts pour la modélisation d'habitat aviaire. L'algorithme évite la multicolinéarité grâce à l'importance de permutation Random Forest et assure la stabilité par répétition bootstrap avec approche par consensus (Meinshausen & Bühlmann, 2010). Il gère efficacement les interactions complexes tout en prévenant le sur-ajustement via la limitation adaptative du nombre de variables.

Spécifiquement en écologie aviaire, VSURF permet la gestion robuste des espèces rares, distingue les variables d'interprétation des variables prédictives, et quantifie l'incertitude via les fréquences de sélection (Genuer et al., 2015).

Limitations et biais

L'algorithme présente une sensibilité excessive aux paramètres, particulièrement problématique pour les espèces rares, et hérite des biais de Random Forest notamment vers les variables continues à forte variabilité (Strobl et al., 2007). La stabilité computationnelle peut être défaillante, et le coût computationnel reste élevé.

Les biais méthodologiques incluent un biais conservateur pouvant sous-estimer l'importance de variables écologiquement pertinentes, et un biais de phase créant une tension entre interprétation et prédiction. Dans le contexte écologique, des biais taxonomiques, spatiaux et temporels peuvent affecter la sélection selon les propriétés statistiques plutôt que la pertinence écologique.

Considérations d'interprétation

La fréquence de sélection ne reflète pas nécessairement l'importance écologique, et les variables sélectionnées sont optimales pour le jeu de données spécifique sans garantie de généralisation. Une validation externe reste indispensable.

4. MÉTHODOLOGIE

4.1 Design d'étude et stratégie d'analyse

Il est important de préciser que la méthodologie présentée ci-dessous a été réalisée sur la base d'une revue de la littérature et d'avis d'experts. Des choix ont été faits pour mettre au point cette modélisation ; ces choix peuvent comporter des biais et ont une importance considérable sur les résultats finaux. Cette méthodologie ne représente qu'une des multiples approches possibles pour l'analyse de ce genre de données et, au vu de la longueur limitée du travail, la méthodologie et les choix sont présentés de manière la plus claire et concise possible, afin de donner une vue d'ensemble et de présenter les étapes clés du modèle.

Voici un diagramme qui reprend l'ensemble des étapes clé du modèle.

PIPELINE MÉTHODOLOGIQUE ADAPTATIF

Modélisation d'habitat aviaire forestier intégrant les données LiDAR

Approche différentielle selon la prévalence spécifique

PHASE I : ACQUISITION ET PRÉTRAITEMENT DES DONNÉES

Acquisition Multi-sources

- Programme SOCWAL
- Variables LiDAR
- Composition forestière
- Métriques paysagères

Classification Taxonomique

- Calcul prévalence par espèce
- Communes : $\geq 15\%$
- Rares : 5-15%
- Très rares : $< 5\%$

Autocorrélation Spatiale

- Analyse variographique
- Extraction portée (range)
- Paramètre défaut : 10 km

PHASE II : VALIDATION CROISÉE SPATIALE ADAPTATIVE

ESPÈCES COMMUNES

Prévalence $\geq 15\%$

- Validation croisée spatiale
- Block Cross-Validation standard
- Stratification équilibrée

ESPÈCES RARES

$\% \leq \text{Prévalence} < 15\%$

- Validation stratifiée spatiale
- Blocks adaptés effectif réduit

ESPÈCES TRÈS RARES

Prévalence $< 5\%$

- Leave-One-Out spatial
- Validation exhaustive (n = observations)

PHASE III : MODÉLISATION COMPARATIVE RANDOM FOREST

Modèle COMPLET

- Variables LiDAR
- Variables de composition forestière
- Variables paysagères

Modèle SANS_LIDAR

- Variables de composition forestière
- Variables paysagères
- Exclusion : LiDAR

Modèle LIDAR_SEUL

- LiDAR uniquement
- Tester la suffisance des données LiDAR seules pour la modélisation

PHASE IV : ÉVALUATION STATISTIQUE ET DIAGNOSTICS

SYNTHÈSE ET INTERPRÉTATION ÉCOLOGIQUE

- Quantification apport LiDAR : 57% espèces bénéficiaires (gain ΔAUC médian +0.016)
- Validation hypothèses : H1 partielle, H2 partielle, H3 invalidée

Figure 1 : Diagramme schématisant de manière synthétique les étapes clé de la modélisation

4.1.1 Approche générale

Cette étude adopte une approche de modélisation d'habitat par apprentissage automatique supervisé pour quantifier l'apport des variables LiDAR dans la prédiction de la distribution des espèces aviaires (Austin, 2007). L'analyse repose sur un design comparatif opposant des modèles intégrant l'ensemble des variables environnementales (incluant les métriques LiDAR) à des modèles excluant spécifiquement les variables tridimensionnelles, permettant ainsi de quantifier précisément leur contribution prédictive.

4.1.2 Pipeline adaptatif selon la prévalence

L'approche méthodologique développée s'adapte automatiquement aux caractéristiques écologiques et statistiques de chaque espèce étudiée. Cette stratification repose sur la prévalence (proportion de présences dans l'échantillon), paramètre déterminant pour l'optimisation des méthodes d'analyse selon les recommandations établies par Fielding & Bell (1997) et Lomba et al. (2010) pour la modélisation d'espèces rares.

Seuils de classification adoptés :

- **Espèces communes** : prévalence $\geq 15\%$
- **Espèces rares** : $5\% \leq$ prévalence $< 15\%$
- **Espèces très rares** : prévalence $< 5\%$

Ces seuils s'appuient sur les travaux de Japkowicz & Stephen (2002) démontrant que le déséquilibre des classes devient critique en dessous de 15% de prévalence, et sur Breiner et al. (2015) établissant qu'en dessous de 5%, des méthodes spécialisées sont nécessaires pour maintenir la robustesse statistique.

4.2 Données environnementales et prétraitement

4.2.1 Calcul des variables

Les variables, préalablement calculées, sont brièvement présentées dans cette section.

Variables LiDAR (n = 28) :

Les métriques LiDAR ont été calculées sur des parcelles circulaires de 300 m de rayon (28,3 ha) à partir des données aéroportées de la campagne wallonne 2020-2021. Ces variables se répartissent selon quatre dimensions fonctionnelles avec des résolutions de calcul spécifiques :

Métriques de hauteur - Résolution : points natifs

- **Hauteur_Max** : hauteur maximale de la végétation
- **Hauteur_Moyenne** : hauteur moyenne de la végétation
- **Hauteur_P95** : 95e percentile des hauteurs
- **Height_Median_Energy_m** : médiane énergétique des hauteurs
- **CHM_Height_StdDev_m** : écart-type des hauteurs du CHM (raster 1m)
- **CHM_Height_CoeffVar_pct** : coefficient de variation des hauteurs (raster 1m)
- **CHM_Height_Range_m** : étendue des hauteurs du CHM (raster 1m)
- **HIQR** : écart interquartile des hauteurs

Métriques de densité - Résolution : points natifs + strates

- **LAD_Mean, LAD_Max, LAD_SD** : densité foliaire moyenne, maximale et écart-type
- **Return_Density_Understory_per_m2** : densité de retours sous-bois (0.5-5m)
- **Return_Density_Midstory_per_m2** : densité de retours étage moyen (5-15m)
- **Return_Density_Canopy_per_m2** : densité de retours canopée (15-60m)
- **Vegetation_Density_Ratio** : ratio densité végétation/sol
- **LiDAR_Penetration_Ratio** : coefficient de pénétration LiDAR
- **CRR** : ratio de relief de canopée
- **Understory_Structural_Complexity** : complexité structurale sous-bois
- **Forest_Edge_Density_per_m2** : densité de lisière (raster 1m)

Métriques de complexité structurale - Résolution : mixte

- **Rumple_Index** : indice de rugosité 3D (raster 1m)
- **Foliage_Height_Diversity_Shannon** : diversité verticale de Shannon (points natifs)
- **Vertical_Complexity_Index** : indice de complexité verticale (10 strates)
- **CHM_Height_StdDev_m** : hétérogénéité structurale (raster 1m)

Métriques de couverture - Résolution : raster 1m

- **Gap_Fraction_Ratio** : fraction de gaps (< 2m)

- **Canopy_Openness_pct** : ouverture de canopée (%)
- **Canopy_Cover_Understory_pct** : couverture sous-bois (0.5-5m)
- **Canopy_Cover_Midstory_pct** : couverture étage moyen (5-15m)
- **Canopy_Cover_Canopy_pct** : couverture canopée (15-50m)

Ces métriques intègrent les trois composantes essentielles de l'habitat aviaire forestier : la stratification verticale, la disponibilité en ressources et l'hétérogénéité spatiale, permettant une caractérisation fine des niches écologiques tridimensionnelles.

Variables de composition forestière (n = 18) :

Dérivées de la cartographie des essences de Bolyn et al. (2022) à 10m de résolution, sur des parcelles circulaires de 300 m de rayon, incluant les proportions d'essences dominantes (**Chenes, Hetres, Epiceas, Douglas, Pins, Melezes, Bouleaux, Peupliers, Autres, Non.ligneux**), les indices de diversité spécifique (Shannon, Simpson) et les variables de structure forestière.

Indices de diversité spécifique

- **Richesse_Essences** : nombre d'essences présentes
- **Indice_Diversite_Shannon**
- **Indice_Diversite_Simpson**
- **Indice_Diversite_Simpson_Modifie**
- **Indice_Uniformite_Shannon**
- **Indice_Uniformite_Simpson**

Variables de structure forestière - Résolution : calculés sur parcelle 300m

- **Pourcentage_Essence_Dominante** : pourcentage de l'essence la plus abondante
- **Coefficient_Variation_Heterogeneite** : coefficient de variation des proportions d'essences

Variables paysagères (n = 32) :

Les variables paysagères ont été calculées à partir de la cartographie de l'utilisation du sol de Wallonie (WAL_UTS_2018) sur des parcelles circulaires de 300 m de rayon. Ces variables se répartissent selon quatre dimensions fonctionnelles :

Variables de couverture d'habitat - Résolution : polygones 1m agrégés sur parcelle 300m

- Couverture_Foret_Sylviculture : pourcentage de forêts de production
- Couverture_Foret_Sapins : pourcentage de sapinières de Noël
- Couverture_Forestiere_Totale : pourcentage forestier total
- Couverture_Agriculture_Intensive : pourcentage d'agriculture intensive
- Couverture_Agriculture_Extensive : pourcentage d'agriculture extensive
- Couverture_Agriculture_Totale : pourcentage agricole total
- Couverture_Prairie : pourcentage de prairies permanentes
- Couverture_Zones_Naturelles_Totale : pourcentage de zones naturelles
- Couverture_Naturelle_Totale : pourcentage total d'habitats naturels
- Couverture_Residentiel_Tres_Faible : pourcentage résidentiel très faible densité
- Couverture_Residentiel_Faible : pourcentage résidentiel faible densité
- Couverture_Residentiel_Moyenne : pourcentage résidentiel densité moyenne
- Couverture_Residentiel_Forte : pourcentage résidentiel forte densité
- Couverture_Residentiel_Totale : pourcentage résidentiel total

Variables d'anthropisation

- Couverture_Anthropique_Total : pourcentage d'infrastructures (transport, industrie, services)
- Couverture_Usage_Inconnu_Total : pourcentage d'usages non définis
- Couverture_Zone_Abandonnee : pourcentage de zones abandonnées
- Indice_Artificialisation : gradient d'artificialisation

Indices de configuration paysagère

- Indice_Division_Paysage : degré de division spatiale
- Indice_Maille_Effective : taille effective des mailles
- Indice_Fragmentation_Split : indice de fragmentation
- Indice_Cohesion_Paysage : cohésion spatiale des habitats
- Indice_Agregation_Paysage : agrégation des éléments paysagers
- Indice_Forme_Paysage : complexité des formes
- Fragmentation_Totale : nombre total de fragments
- Richesse_Habitat : nombre de types d'habitats présents

Indices synthétiques d'habitat - Résolution : calculées sur parcelle 300m

- Shannon_Habitat : diversité de Shannon des habitats

- **Type_Dominant** : type d'habitat dominant
- **Pourcentage_Dominant** : pourcentage du type dominant
- **Indice_Naturalite** : gradient de naturalité
- **Indice_Habitat_Forestier** : spécialisation forestière
- **Indice_Mosaïque_Agricole** : complexité agricole
- **Indice_Pression_Urbaine** : intensité de pression urbaine

4.2.2 Stratégie de génération des pseudo-absences

La génération des pseudo-absences sur le jeu de données aviaires, suit une approche d'échantillonnage aléatoire stratifié environnementalement pour minimiser les biais géographiques (Barbet-Massin et al., 2012).

Paramètres d'échantillonnage :

- **Espèces communes (1:1)** : Autant de pseudo-absences que de présences
- **Espèces rares (1:2)** : Deux fois plus de pseudo-absences que de présences. Compenser le déséquilibre naturel des espèces rares
- **Distance minimale aux présences** : 500m (supérieure à la précision GPS)

4.3 Caractérisation de l'autocorrélation spatiale

4.3.1 Diagnostic de l'autocorrélation

L'autocorrélation spatiale a été quantifiée par analyse variographique suivant les principes établis par Matheron (1963). Pour chaque espèce, un variogramme empirique a été calculé sur les résidus d'un modèle null (présence ~ 1) selon la formule :

$$\gamma(h) = 1/(2N(h)) \sum [Z(x_i) - Z(x_i+h)]^2$$

où $\gamma(h)$ est la semi-variance à la distance h , $N(h)$ le nombre de paires de points séparés par la distance h , et $Z(x_i)$ la valeur résiduelle au point x_i .

Ajustement du modèle théorique : Un modèle sphérique a été ajusté par moindres carrés pondérés :

$$\gamma(h) = C_0 + C \times [1.5(h/a) - 0.5(h/a)^3] \text{ pour } h \leq a \quad \gamma(h) = C_0 + C \text{ pour } h > a$$

où C_0 est l'effet de pépité, C la variance structurelle, et a la portée (range).

4.3.2 Paramètres d'autocorrélation extraits

Pour chaque espèce, trois paramètres clés ont été extraits selon les standards géostatistiques écologiques trouvés (Webster & Oliver, 2007) :

- **Portée (range)** : distance au-delà de laquelle l'autocorrélation devient négligeable
- **Effet de pépité relatif** : $C_0/(C_0+C)$, indicateur de l'autocorrélation à courte distance
- **Variance structurelle** : C , mesure de la variabilité spatiale organisée

Ces paramètres informent le dimensionnement des blocs spatiaux pour la validation croisée selon les recommandations de Dormann et al. (2007). En cas d'échec de l'ajustement variographique, une portée par défaut de 10 km est appliquée, correspondant à l'ordre de grandeur typique de l'autocorrélation spatiale pour les données d'occurrence d'oiseaux forestiers (Dormann et al., 2007).

4.4 Validation croisée spatiale adaptative

4.4.1 Principe de l'approche adaptative

La stratégie de validation croisée s'adapte automatiquement à la prévalence de l'espèce pour optimiser le compromis entre indépendance spatiale, représentativité statistique et stabilité des estimateurs (Valavi et al., 2019). Cette approche répond aux défis spécifiques de la modélisation d'espèces présentant des prévalences et des patterns spatiaux très variables.



Figure 2 : Exemple de folds/blocs spatiaux générés par la validation croisée spatiale adaptative selon la méthode BlockCV

4.4.2 Méthode 1 : Block Cross-Validation standard (espèces communes \geq 15%)

Pour les espèces communes, le nombre élevé de présences permet l'application de méthodes standard sans risque majeur de déséquilibre des blocs (Hastie et al., 2009).

Paramètres :

- Nombre de blocs : 10 (compromis optimal biais-variance selon Kohavi, 1995)
- Taille des blocs : $\max(2 \times \text{portée d'autocorrélation}, 5000\text{m})$
- Sélection : assignation aléatoire des blocs aux folds
- Optimisation : 100 itérations pour minimiser la variance inter-blocs

Implémentation : L'algorithme BlockCV est un package dans R qui crée des blocs géographiques contigus selon une grille spatiale, puis assigne aléatoirement ces blocs aux différents folds (blocs de validation croisée). Les détails de cette méthode sont expliqués dans l'article (Valavi et al., 2019).

Les BlockCV permettent une évaluation plus réaliste du modèle en tenant compte de l'autocorrélation spatiale ou environnementale. Ils réduisent le risque de surévaluation des performances en évitant que des points proches soient répartis dans des blocs différents. De plus, ces méthodes sont faciles à visualiser et peuvent être adaptées à divers types de données spatiales ou environnementales.

Cependant, il faut garder à l'esprit que :

- La qualité des résultats dépend fortement du choix de la taille des blocs ;
- La méthode peut être complexe à paramétrer et à mettre en œuvre, nécessitant une compréhension approfondie des données d'entrée ;
- Elle ne résout pas totalement les problèmes d'autocorrélation, notamment si la structure spatiale est très complexe ou hétérogène, ce qui est le cas pour certaines données de cette études.

4.4.3 Méthode 2 : Validation stratifiée spatiale (espèces rares 5-15%)

Les espèces rares nécessitent une stratification explicite pour garantir la présence d'au moins 3 occurrences par fold et maintenir la cohérence spatiale (Breiner et al., 2015).

Paramètres :

- Nombre de folds : $\min(5, \max(3, \lfloor n_{\text{présences}}/2 \rfloor))$
- Minimum de présences par fold : 3
- Clustering : k-means spatial sur les coordonnées des présences (Fortin & Dale, 2005)
- Attribution des absences : assignation au fold de présences le plus proche

Procédure :

1. **Clustering spatial des présences** : Application d'un k-means géographique pour créer des groupes spatialement cohérents
2. **Validation des contraintes** : Vérification que chaque fold contient ≥ 3 présences

3. **Attribution des absences** : Chaque absence est assignée au fold du point de présence le plus proche selon la distance euclidienne

4.4.4 Méthode 3 : Leave-One-Out spatial (espèces très rares < 5%)

Pour les espèces très rares, l'approche Leave-One-Out maximise l'utilisation des rares données disponibles tout en maintenant l'indépendance spatiale (Lomba et al., 2010).

Paramètres :

- Nombre de folds : nombre de présences
- Distance d'exclusion : portée d'autocorrélation spatiale
- Buffer d'exclusion : tous les points dans un rayon = portée autour de chaque présence test

Procédure :

1. **Définition des folds individuels** : Chaque présence constitue un fold de test unique
2. **Exclusion spatiale** : Tous les points dans un buffer de rayon = portée d'autocorrélation sont exclus de l'entraînement (Telford & Birks, 2005)
3. **Validation de l'indépendance** : Contrôle que la distance entre points d'entraînement et de test > portée

4.4.5 Évaluation de la qualité des folds

Métriques de qualité calculées :

- Proportion de folds utilisables (≥ 1 présence)
- Coefficient de variation de la taille des folds : $CV = (\sigma/\mu) \times 100$
- Distance minimale inter-folds
- Prévalence moyenne par fold et écart-type

Critères de classification :

- **Excellente** : $\geq 90\%$ folds utilisables, $CV < 25\%$
- **Bonne** : 70-89% folds utilisables, $CV < 40\%$
- **Acceptable** : 50-69% folds utilisables, $CV < 60\%$
- **Problématique** : $< 50\%$ folds utilisables ou $CV > 60\%$

Ces critères sont adaptés de Hastie et al. (2009) pour l'évaluation de la qualité des partitions en validation croisée.

4.5 Sélection de variables par VSURF

4.5.1 Principe de l'algorithme VSURF

L'algorithme VSURF (Variable Selection Using Random Forests) développé par Genuer et al. (2015) combine l'efficacité de Random Forest pour l'estimation d'importance avec une procédure de sélection en trois étapes : élimination des variables non-informatives, sélection pour l'interprétation, et sélection pour la prédiction.

4.5.2 Configuration adaptative selon la prévalence

Les espèces rares nécessitent une approche plus conservatrice pour compenser la variabilité accrue des estimateurs d'importance due au faible échantillonnage (Breiner et al., 2015).

Tableau 1: Paramètres de configuration VSURF adaptatifs selon la catégorie de prévalence des espèces

Catégorie	Seuil consensus	nombre d'arbres	Bootstrap	Justification
Communes ($\geq 15\%$)	40%	1000	100	Mode relaxed : échantillon suffisant pour stabilité
Rares (5-15%)	60%	1500	200	Mode conservateur : réduction de la variabilité
Très rares ($< 5\%$)	80%	2000	500	Mode ultra-conservateur : maximisation de la robustesse

Critères d'arrêt et de convergence :

- Stabilité : coefficient de variation des importances $< 10\%$
- Convergence : différence entre itérations consécutives $< 1\%$
- Maximum : 1000 itérations par étape

4.5.3 Validation de la stabilité des sélections

Pour chaque espèce, la stabilité de la sélection a été évaluée par bootstrap ($n = 500$) selon la méthode de Meinshausen & Bühlmann (2010). La probabilité de sélection de chaque variable est calculée, et seules les variables avec une probabilité $>$ seuil de consensus sont retenues.

4.6 Modélisation par Random Forest

4.6.1 Justification du choix algorithmique

Random Forest a été sélectionné pour ses propriétés optimales dans le contexte de cette étude (Breiman, 2001) :

- **Robustesse aux variables corrélées** : approprié pour les métriques LiDAR inter-corrélées
- **Gestion native du déséquilibre** : pondération automatique des classes
- **Estimation d'importance** : mesure MDA (Mean Decrease Accuracy) non-biaisée
- **Résistance au sur-ajustement** : particulièrement critique pour les espèces rares

4.6.2 Paramétrage et optimisation

Hyperparamètres fixés :

- Nombre d'arbres (ntree) : 1000 (compromis performance-temps de calcul)
- Variables par nœud (mtry) : \sqrt{p} (p = nombre de variables, optimum théorique selon Breiman, 2001)
- Taille minimale des nœuds terminaux : 1 (maximisation de la granularité)
- Échantillonnage : bootstrap avec remplacement

Gestion du déséquilibre des classes : Pondération inverse de la prévalence appliquée selon la formule de Japkowicz & Stephen (2002) : $w_1 = n/(2 \times n_1)$ et $w_0 = n/(2 \times n_0)$

où n est le nombre total d'observations, n_1 le nombre de présences, et n_0 le nombre d'absences.

4.6.3 Procédure de validation croisée intégrée

Pour chaque fold de validation :

1. **Entraînement** : modèle Random Forest sur (k-1) folds avec variables sélectionnées par VSURF
2. **Prédiction** : probabilités de présence sur le fold de test
3. **Évaluation** : calcul des métriques de performance
4. **Agrégation** : moyenne et écart-type des performances sur tous les folds

4.7 Métriques d'évaluation des performances

4.7.1 Métriques primaires

AUC-ROC (Area Under the ROC Curve) : Métrique principale pour l'évaluation de la capacité discriminante, calculée selon :

$$AUC = \int_0^1 TPR(FPR^{-1}(t))dt$$

où TPR est le taux de vrais positifs et FPR le taux de faux positifs.

Seuils d'interprétation adoptés (Swets, 1988) :

- Excellente : $AUC \geq 0.9$
- Bonne : $0.8 \leq AUC < 0.9$
- Acceptable : $0.7 \leq AUC < 0.8$
- Faible : $AUC < 0.7$

4.7.2 Métriques complémentaires

AUC-PR (Precision-Recall) : Alternative à l'AUC-ROC pour les classes très déséquilibrées, calculée sur la courbe Précision-Rappel selon les recommandations de Saito & Rehmsmeier (2015).

4.8 Analyse comparative de l'apport LiDAR

4.8.1 Design comparatif

L'apport des variables LiDAR est quantifié par comparaison directe de trois configurations :

1. **Modèle COMPLET** : toutes les variables (LiDAR + composition + paysage)

2. **Modèle SANS_LIDAR** : variables de composition et paysagères uniquement
3. **Modèle LIDAR_SEUL** : variables LiDAR uniquement (analyse de suffisance)

4.8.2 Tests de significativité statistique

Test de DeLong : Comparaison statistique des AUC-ROC selon la méthode de DeLong et al. (1988) :

$$z = (AUC_1 - AUC_2) / \sqrt{(SE_1^2 + SE_2^2 - 2 \times Cov_{12})}$$

avec correction de Bonferroni pour comparaisons multiples.

Bootstrap des différences : Estimation par bootstrap (n = 1000) des intervalles de confiance à 95% pour les différences d'AUC.

4.8.3 Classification de l'apport LiDAR

Seuils de classification quantitative :

- **Gain substantiel** : $\Delta AUC \geq 0.05$ et $p < 0.01$
- **Gain modéré** : $0.02 \leq \Delta AUC < 0.05$ et $p < 0.05$
- **Gain marginal** : $0.01 \leq \Delta AUC < 0.02$ et $p < 0.1$
- **Aucun gain** : $\Delta AUC < 0.01$ ou $p \geq 0.1$

4.9 Analyse de sensibilité et validation

4.9.1 Tests de robustesse

Sensibilité aux paramètres spatiaux testée ultérieurement :

- Variation de la taille des blocs : $\times 0.5$, $\times 1.5$, $\times 2.0$
- Impact du ratio présences/absences : 1:1, 1:2, 1:3
- Robustesse aux seuils VSURF : $\pm 10\%$ sur les seuils de consensus

4.9.2 Analyse des résidus

Diagnostic spatial des résidus :

- Test de Moran sur les résidus de prédiction pour détecter l'autocorrélation résiduelle (Dormann et al., 2007)
- Analyse des patterns de sur/sous-prédiction
- Validation de l'absence d'autocorrélation résiduelle selon les critères de Telford & Birks (2005)

4.10 Environnement computationnel et reproductibilité

4.10.1 Spécifications techniques

Environnement logiciel :

- Langage : R version 4.3.2
- Packages principaux : randomForest v4.7-1, VSURF v1.2.0, blockCV v3.1-4, sf v1.0-14

Contrôles de reproductibilité :

- Graine aléatoire fixée : set.seed(123) pour tous processus stochastiques
- Versions des packages documentées
- Validation croisée des résultats sur échantillon test

4.11 Limitations méthodologiques

4.11.1 Limitations inhérentes au design

Données de présence-absence :

- Biais de détection non corrigés (différences d'observabilité inter-espèces)
- Pseudo-absences générées plutôt qu'observées

Contraintes spatiales :

- Design en chaînes de points d'écoute non-optimal pour l'indépendance spatiale
- Autocorrélation résiduelle possible malgré les précautions
- Couverture géographique limitée à la Wallonie et agrégée

4.11.2 Limitations algorithmiques

Random Forest :

- Biais potentiel vers les variables continues (Strobl et al., 2007)
- Interactions d'ordre supérieur non explicitement modélisées
- Sensibilité aux variables corrélées malgré la robustesse générale

VSURF :

- Stabilité des sélections dépendante de la taille d'échantillon (Genuer et al., 2015)
- Seuils de consensus arbitraires malgré les justifications empiriques
- Possible sur-sélection pour les espèces communes

4.11.3 Considérations écologiques

Échelles temporelles :

- Données LiDAR et ornithologiques non strictement contemporaines
- Dynamique forestière non prise en compte
- Variations saisonnières de l'utilisation de l'habitat non intégrées

Variables manquantes :

- Ressources alimentaires non quantifiées directement
- Prédation et interactions interspécifiques non modélisées
- Microhabitat et variables à très fine échelle limitées

4.11.4 Limitations des tests statistiques

Une critique de la méthodologie des tests statistiques se trouve en annexe E de ce travail.

5. RÉSULTATS

Pour des raisons de longueur limitée de ce document, cette section présente une vue d'ensemble des résultats et performances du modèle à l'échelle de l'ensemble des espèces analysées. Les résultats détaillés par espèce, incluant les variables sélectionnées, les directions d'effet, les métriques de performance et les analyses de diagnostic, sont présentés de manière plus lisible et interactive sur la plateforme dédiée accessible via ce lien :

<https://sites.google.com/view/habitatsaviaires/accueil>

Le choix de cette présentation duale répond à un objectif d'accessibilité : permettre aux naturalistes, gestionnaires forestiers et acteurs de la conservation d'accéder facilement et directement aux informations pertinentes pour chaque espèce d'intérêt, sans diluer l'information concrète dans un rapport technique dense. La plateforme interactive propose des visualisations et un outil de filtrage permettant une exploration ciblée des résultats selon les besoins spécifiques de chaque utilisateur. Cette plateforme a été créée par intelligence artificielle et n'est qu'un prototype qui pourrait être étoffé et amélioré dans le futur.

5.1 Caractérisation des espèces étudiées

Distribution des prévalences et classification adaptative

L'analyse porte sur 61 espèces aviaires dont la modélisation a abouti avec succès. Les deux échecs concernent le Pouillot fitis et la Pie-grièche écorcheur, pour lesquels des erreurs techniques lors du processus de modélisation n'ont pas permis d'obtenir des résultats exploitables.

Répartition par catégorie de prévalence :

- **Espèces communes** (prévalence $\geq 15\%$) : 28 espèces (45,9%)
- **Espèces rares** ($5\% \leq$ prévalence $< 15\%$) : 19 espèces (31,1%)
- **Espèces très rares** (prévalence $< 5\%$) : 14 espèces (23,0%)

Les espèces les plus communes incluent des généralistes bien connus comme le Pigeon ramier (84,3% de prévalence), la Fauvette à tête noire (87,9%) et le Pouillot véloce (80,9%), tandis que les espèces très rares comprennent principalement des spécialistes tels que le Pic noir (6,1%), la Bergeronnette flavéole (4,2%) et le Vanneau huppé (2,8%).

Autocorrélation spatiale et qualité de la validation croisée

L'analyse variographique révèle une forte hétérogénéité des patterns d'autocorrélation spatiale entre espèces. La portée d'autocorrélation varie de 5 km (espèces forestières spécialisées) à plus de 30 km (espèces ubiquistes), avec une médiane à 10 km. Cette variabilité reflète les différences d'écologie spatiale et de mobilité entre espèces, justifiant l'approche adaptative de dimensionnement des blocs spatiaux.

Tableau 2 : Qualité des folds de validation croisée spatiale

Qualité des folds	Critères	Nombre d'espèces	Pourcentage	Implications
Excellente	≥90% folds utilisables, CV<25%	25	41%	Résultats très fiables
Bonne	70-89% folds utilisables, CV<40%	17	28%	Résultats fiables
Acceptable	50-69% folds utilisables, CV<60%	12	20%	Résultats à nuancer
Problématique	<50% folds utilisables ou CV>60%	7	11%	Interprétation prudente

Les espèces à qualité problématique correspondent principalement aux espèces très rares ou présentant une distribution spatiale extrêmement agrégée, comme le Héron cendré ou certaines espèces urbaines. Cette limitation, inhérente à la structure en chaînes des données SOCWAL, constitue un biais méthodologique important mais ne remet pas en cause la validité générale des résultats pour la majorité des espèces.

Biais spatial d'échantillonnage : Il convient de noter que la structure en chaînes de points d'écoute SOCWAL, privilégiant les zones accessibles, génère potentiellement une sur-représentation des habitats de lisière au détriment des zones de cœur de massif forestier. Cette configuration spatiale non-aléatoire peut biaiser l'importance apparente de certaines variables, particulièrement les métriques liées aux effets de bordure, sans que cela reflète nécessairement leur importance écologique réelle pour l'ensemble de l'avifaune forestière.

5.2 Performance globale du modèle et apport des données LiDAR

Le modèle Random Forest a produit des performances variables selon les espèces, avec des AUC-ROC s'échelonnant de 0.497 Corbeau freux (*Corvus frugilegus*) à 0.986 rousserolle verderolle (*Acrocephalus palustris*). La distribution des performances révèle :

- **Performance excellente** ($AUC \geq 0.90$) : 4 espèces (6,6%)
- **Performance bonne** ($0.80 \leq AUC < 0.90$) : 17 espèces (27,9%)
- **Performance acceptable** ($0.70 \leq AUC < 0.80$) : 25 espèces (41,0%)
- **Performance faible** ($AUC < 0.70$) : 15 espèces (24,6%)

Ces résultats sont cohérents avec la littérature en modélisation d'habitat aviaire, où les performances dépendent fortement de la spécialisation écologique des espèces et de la qualité des données disponibles.

Apport quantitatif des variables LiDAR

L'analyse comparative révèle un gain médian d'AUC-ROC de +0.016 lors de l'intégration des variables LiDAR (différence entre modèles COMPLET et SANS_LIDAR). Cependant, cette valeur masque une importante variabilité inter-espèces :

Distribution des gains LiDAR :

- **Gain substantiel** (≥ 0.05) : 8 espèces (13,1%)
- **Gain modéré** (0.02-0.05) : 12 espèces (19,7%)
- **Gain marginal** (0.01-0.02) : 15 espèces (24,6%)
- **Aucun gain** ou gain négatif : 26 espèces (42,6%)

Espèces avec les gains LiDAR les plus substantiels :

1. Coucou gris : +0.0556
2. Fauvette des jardins : +0.0382
3. Tarier pâtre : +0.0328
4. Buse variable : +0.0322
5. Fauvette grisette : +0.0243

Ces résultats nuancent fortement l'hypothèse H1 selon laquelle les variables LiDAR amélioreraient systématiquement la prédiction d'habitat. Pour près de 43% des espèces analysées, l'ajout des variables LiDAR n'apporte aucun bénéfice mesurable, voire dégrade

légèrement les performances. Cette proportion importante d'espèces non-bénéficiaires suggère que la valeur ajoutée du LiDAR est hautement spécifique aux exigences écologiques particulières de chaque espèce.

Variabilité selon la spécialisation écologique

L'analyse de l'hypothèse H2 révèle des patterns complexes qui ne suivent pas strictement le gradient généraliste-spécialiste attendu. Contrairement aux prédictions théoriques, plusieurs espèces reconnues comme spécialistes ne bénéficient pas significativement des variables LiDAR, tandis que certaines espèces à écologie plus flexible montrent des gains substantiels.

Exemples contrastés :

- Le **Pic noir** (spécialiste forestier emblématique) : gain LiDAR négligeable (-0.007)
- La **Sittelle torchepot** (spécialiste des vieilles forêts) : gain minimal (+0.005)
- Le **Coucou gris** (parasite généraliste) : gain majeur (+0.0556)

Cette discordance suggère que soit la complexité structurelle 3D n'est pas universellement déterminante pour tous les spécialistes forestiers, soit que les métriques LiDAR utilisées ne traduisent pas complètement la complexité structurelle réellement perçue et utilisée par ces espèces. Cette observation remet en question les présupposés théoriques sur lesquels reposait l'hypothèse H2, tout en soulignant les limites potentielles des variables LiDAR utilisées dans cette étude pour caractériser l'aspects de l'habitat forestier.

Variables LiDAR universellement importantes

Contrairement à l'hypothèse H3 qui prédisait la prédominance des variables de complexité structurelle, les métriques de hauteur et de couverture dominant les sélections :

Top 5 des variables LiDAR les plus fréquemment sélectionnées :

1. **CHM_Height_StdDev_m** (écart-type des hauteurs) : 67% des espèces
2. **Canopy_Cover_Canopy_pct** (couverture de canopée) : 61% des espèces
3. **Forest_Edge_Density_per_m2** (densité de lisière) : 59% des espèces
4. **Return_Density_Canopy_per_m2** (densité de retours canopée) : 48% des espèces
5. **LiDAR_Penetration_Ratio** (ratio de pénétration) : 43% des espèces

L'écart-type des hauteurs (CHM_Height_StdDev_m) émerge comme la variable LiDAR la plus universellement informative, présente dans le modèle de plus de deux tiers des espèces. Cette métrique caractérise l'hétérogénéité structurelle verticale, mais de manière plus simple que les indices de complexité sophistiqués initialement privilégiés.

Les variables de complexité pure (Rumple Index, diversité de hauteur du feuillage) ne figurent que rarement dans le top des sélections, contredisant partiellement l'hypothèse H3. Cette observation suggère que les oiseaux répondent davantage à des caractéristiques structurelles simples et robustes qu'à des indices de complexité mathématiquement élaborés.

Considération critique sur Forest_Edge_Density_per_m2 : La forte présence de cette variable dans 59% des résultats doit être interprétée avec prudence. Cette prédominance pourrait refléter un biais d'échantillonnage spatial plutôt qu'une réelle importance écologique universelle. En effet, les points d'écoute SOCWAL, organisés en chaînes accessibles, sont potentiellement concentrés en zones de lisière plutôt qu'en plein cœur des massifs forestiers. Cette distribution spatiale non-aléatoire pourrait artificiellement amplifier l'importance apparente des variables de lisière dans le modèle, sans que cela reflète nécessairement un facteur écologique déterminant pour l'ensemble des espèces. Cette limitation souligne l'importance de distinguer les patterns statistiques des mécanismes écologiques authentiques dans l'interprétation des résultats.

Tableau 3 : Fréquence de sélection des 20 variables environnementales les plus utilisées dans la modélisation de l'habitat de 61 espèces aviaires en Wallonie

	Variable	Fréquence	% d'utilisation
1	CHM_Height_StdDev_m	41	67.2 %
2	Canopy_Cover_Canopy_pct	37	60.7 %
3	Forest_Edge_Density_per_m2	36	59.0 %
4	Return_Density_Canopy_per_m2	29	47.5 %
5	LiDAR_Penetration_Ratio	26	42.6 %
6	CHM_Height_CoeffVar_pct	25	41.0 %
7	Couverture_Agriculture_Totale	24	39.3 %
8	Canopy_Cover_Midstory_pct	23	37.7 %
9	Hauteur_P95	22	36.1 %
10	Couverture_Prairie	21	34.4 %
11	Chenes	20	32.8 %
12	Couverture_Forestiere_Totale	19	31.1 %
13	Height_Median_Energy_m	18	29.5 %
14	Couverture_Agriculture_Intensive	17	27.9 %
15	Vegetation_Density_Ratio	16	26.2 %
16	Hauteur_Moyenne	15	24.6 %
17	Gap_Fraction_Ratio	14	23.0 %
18	Couverture_Anthropique_Total	13	21.3 %
19	Return_Density_Midstory_per_m2	12	19.7 %
20	Canopy_Cover_Understory_pct	11	18.0 %

Le tableau ci-dessus révèle une domination marquée des variables dérivées du LiDAR dans la sélection automatique de variables, avec 14 des 20 variables les plus fréquemment retenues (70%). Cette prédominance souligne l'importance cruciale de la structure tridimensionnelle des habitats forestiers pour expliquer la distribution des espèces aviaires wallonnes.

Patterns de sélection par guildes écologique

L'analyse des patterns de sélection révèle des préférences variables selon les guildes fonctionnelles :

Spécialistes de canopée (Sittelle torchepot, Grimpereau des jardins) :

- Prédominance des variables de hauteur (Height_Median_Energy_m, Hauteur_P95)

- Importance des métriques de densité canopée (Return_Density_Canopy_per_m2)

Espèces de sous-bois (Rouge-gorge, Trogodyte mignon) :

- Sélection fréquente des variables de pénétration (LiDAR_Penetration_Ratio)
- Importance des métriques de couverture intermédiaire (Canopy_Cover_Midstory_pct)

Espèces d'habitats ouverts (Alouette des champs, Bruant jaune) :

- Variables d'ouverture privilégiées (Gap_Fraction_Ratio, Canopy_Openness_pct)
- Corrélations négatives avec les métriques de densité forestière

Cette différenciation par guildes, bien que moins marquée qu'attendu, supporte partiellement l'hypothèse H2 sur la variabilité selon la spécialisation écologique.

5.4 Analyse critique des limitations méthodologiques

Biais inhérents au design spatial

L'organisation des données SOCWAL en chaînes de points d'écoute génère, pour certaines espèces, des patterns d'autocorrélation spatiale très marqués. Sept espèces présentent des portées d'autocorrélation supérieures à 30 km, ce qui dépasse largement les capacités de gestion offertes par la validation croisée spatiale mise en place. Cette limitation structurelle affecte particulièrement les espèces à distribution fortement agrégée, ainsi que certaines espèces urbaines présentant un pattern spatial discontinu.

Pour ces espèces, les résultats doivent être interprétés avec une extrême prudence, car la validation croisée spatiale ne parvient pas à garantir l'indépendance statistique requise.

Variabilité des configurations VSURF

Bien que l'approche adaptative ait été conçue pour optimiser la robustesse selon la prévalence, certaines configurations demeurent suboptimales :

- Mode "ultra-conservateur" (espèces très rares) : risque de sous-sélection par excès de parcimonie
- Mode "relaxed" (espèces communes) : possible sur-sélection malgré les contraintes

La stabilité de sélection décroissante avec la rareté ($0.87 \rightarrow 0.64$) témoigne des limites intrinsèques de l'approche pour les espèces les plus contraintes en données. Cette dégradation est statistiquement inévitable mais limite la fiabilité des résultats pour les espèces de plus grand intérêt conservatoire.

Espèces problématiques identifiées

14% des espèces analysées (n=9) présentent des résultats de qualité insuffisante pour une interprétation écologique robuste :

Espèces à performances aberrantes :

- Corbeau freux (*Corvus frugilegus*) — AUC = 0.497 : sous-échantillonnage critique
- Pigeon colombin (*Columba oenas*) — AUC = 0.521 : distribution spatiale problématique
- Chardonneret élégant (*Carduelis carduelis*) — AUC = 0.531 : probable déséquilibre dans les données

Espèces à qualité de folds problématique :

- Héron cendré, Canard colvert : autocorrélation spatiale trop forte pour le modèle actuel.
- Plusieurs espèces très rares : effectifs insuffisants malgré les adaptations

Ces limitations soulignent que la modélisation automatisée, même adaptative, ne peut compenser des déficiences fondamentales dans la qualité ou la structure des données d'entrée.

5.4 Validation écologique des résultats : cohérence avec les connaissances biologiques

Pour démontrer que malgré les limitations méthodologiques identifiées, le modèle fournit des résultats écologiquement cohérents et donc exploitables, nous présentons ci-dessous l'analyse des variables sélectionnées et de leurs directions pour plusieurs espèces emblématiques, comparées aux connaissances écologiques établies.

Pic noir : spécialiste du bois mort et des vieux peuplements

Variables LiDAR sélectionnées et directions :

- **Rumple_Index** : effet positif fort ($\Delta = +0.237$)
- **Shannon_Habitat** : effet négatif fort ($\Delta = -0.111$)
- **Pourcentage_Dominant** : effet positif fort ($\Delta = +0.111$)

Cohérence écologique : Ces résultats s'accordent parfaitement avec l'écologie connue du Pic noir. Le Pic noir nécessite de grandes superficies boisées (200-500ha) avec des arbres de gros diamètre âgés d'au moins 120 ans, particulièrement les vieux hêtres. L'effet positif du Rumple_Index reflète sa préférence pour les structures forestières complexes où il peut creuser ses loges dans des arbres matures. Le Pic noir creuse de larges loges dans des hêtres âgés de plus de 100 ans, expliquant sa corrélation positive avec les peuplements dominés par une essence (Pourcentage_Dominant élevé), typique des vieilles hêtraies pures. L'effet négatif du Shannon_Habitat confirme sa préférence pour les habitats forestiers homogènes et étendus plutôt que les mosaïques paysagères diversifiées.

Sittelle torchepot : cavicole des vieux chênes

Variables LiDAR sélectionnées et directions :

- **Canopy_Cover_Canopy_pct** : effet positif fort ($\Delta = +0.672$)
- **CHM_Height_StdDev_m** : effet positif fort ($\Delta = +0.653$)
- **Height_Median_Energy_m** : effet positif fort ($\Delta = +0.599$)
- **Chenes** : effet positif fort ($\Delta = +0.588$)

Cohérence écologique : Les résultats correspondent exactement aux exigences écologiques documentées de la Sittelle torchepot. Cette espèce privilégie les vieilles forêts avec de grands et vieux arbres, notamment des chênes, qui lui fournissent des cavités pour nicher. L'effet positif marqué de Canopy_Cover_Canopy_pct et Height_Median_Energy_m reflète sa dépendance aux arbres matures, comme les chênes, les hêtres ou les bouleaux, qui présentent souvent des cavités naturelles. L'importance de CHM_Height_StdDev_m souligne sa préférence pour l'hétérogénéité structurelle qui maximise la disponibilité en sites de nidification et ressources alimentaires dans différentes strates.

Alouette des champs : spécialiste des milieux ouverts

Variables LiDAR sélectionnées et directions :

- **Vegetation_Density_Ratio** : effet négatif fort ($\Delta = -0.424$)
- **Return_Density_Midstory_per_m2** : effet négatif fort ($\Delta = -0.412$)
- **Pourcentage_Essence_Dominante** : effet négatif fort ($\Delta = -0.385$)
- **Indice_Mosaïque_Agricole** : effet positif fort ($\Delta = +0.389$)

Cohérence écologique : Ces résultats illustrent parfaitement l'écologie de l'Alouette des champs, espèce typique des milieux agricoles ouverts. Les effets négatifs forts des variables de densité végétale (Vegetation_Density_Ratio, Return_Density_Midstory_per_m2) confirment son évitement des milieux forestiers denses. L'effet positif de l'Indice_Mosaïque_Agricole correspond à sa préférence pour les paysages agricoles diversifiés où elle niche au sol dans les cultures. L'effet négatif du Pourcentage_Essence_Dominante indique son évitement des zones dominées par des peuplements forestiers homogènes.

Troglodyte mignon : insectivore de sous-bois

Variables LiDAR sélectionnées et directions :

- **Canopy_Cover_Canopy_pct** : effet positif fort ($\Delta = +0.393$)
- **CHM_Height_StdDev_m** : effet positif fort ($\Delta = +0.389$)
- **LiDAR_Penetration_Ratio** : effet négatif fort ($\Delta = -0.351$)
- **Return_Density_Canopy_per_m2** : effet positif fort ($\Delta = +0.366$)

Cohérence écologique : Ces résultats reflètent fidèlement l'écologie du Troglodyte mignon, espèce forestière exploitant principalement le sous-bois dense. L'effet positif de Canopy_Cover_Canopy_pct indique sa préférence pour les forêts à couvert fermé qui maintiennent un microclimat humide favorable à sa recherche d'insectes au sol. L'effet négatif du LiDAR_Penetration_Ratio confirme sa préférence pour les milieux à couvert dense où la pénétration lumineuse est limitée. L'importance de CHM_Height_StdDev_m souligne son besoin d'hétérogénéité verticale créant des microhabitats diversifiés dans le sous-bois.

Validation de la cohérence des patterns inter-espèces

Gradients écologiques cohérents : L'analyse comparative révèle des patterns cohérents le long de gradients écologiques établis :

- **Gradient ouvert-fermé** : Alouette des champs (variables négatives de densité) vs Troglodyte mignon (variables positives de couvert)
- **Gradient structure simple-complexe** : Pic noir (spécialiste peuplements homogènes) vs Sittelle torchepot (hétérogénéité structurelle)
- **Gradient strates** : Variables de canopée pour Sittelle torchepot vs variables de pénétration pour espèces de sous-bois

Robustesse des directions d'effet : La cohérence entre directions d'effet prédites et écologie connue valide la pertinence biologique du modèle, même pour les espèces à performance AUC modérée. Cette observation suggère que les variables sélectionnées captent des signaux écologiques authentiques, même lorsque la capacité prédictive globale est limitée par d'autres facteurs (autocorrélation spatiale, qualité des données).

Cette validation écologique démontre que malgré ses limitations méthodologiques, le pipeline développé produit des résultats biologiquement cohérents et donc exploitables pour la compréhension des relations habitat-espèce et l'orientation des stratégies de conservation.

5.5 Synthèse critique et implications écologiques

Validation nuancée des hypothèses de recherche

Hypothèse H1 (amélioration systématique par les variables LiDAR) :

- Validation pour 57% des espèces avec gain mesurable
- Invalidation pour 43% des espèces sans bénéfice
- **Conclusion** : L'apport LiDAR est **espèce-spécifique** et non universellement bénéfique

Hypothèse H2 (variation selon spécialisation écologique) :

- Patterns différentiels observés entre guildes
- Exceptions notables remettant en question la règle générale
- **Conclusion** : La spécialisation écologique est **prédictive mais non déterminante**

Hypothèse H3 (prédominance des variables de complexité) :

- Variables de hauteur et couverture dominantes sur les indices de complexité
- Métriques simples plus universellement informatives

- **Conclusion : La simplicité structurelle prime sur la complexité mathématique**

Implications pour la gestion forestière

Les résultats identifient **l'hétérogénéité des hauteurs de canopée** comme le déterminant structural le plus universellement important pour l'avifaune forestière wallonne. Cette découverte a des implications pratiques majeures :

Recommandations sylvicoles:

1. **Maintien de l'hétérogénéité verticale** : éviter les peuplements mono-strates
2. **Gestion différenciée des lisières** : importance pour de nombreuses espèces, bien que cette recommandation doive être nuancée par le possible biais d'échantillonnage spatial vers les zones de bordure dans les données SOCWAL
3. **Diversification des couvertures de canopée** : gradients plutôt qu'uniformité

Dans le contexte du Règlement européen sur la restauration de la nature, qui oblige les pays de l'UE à renverser d'ici 2030 la tendance au déclin des indicateurs des oiseaux forestiers, ces résultats fournissent des orientations opérationnelles concrètes pour les gestionnaires forestiers.

Espèces à intérêt conservatoire prioritaire

Parmi les espèces de l'indicateur européen des oiseaux forestiers présentes dans cette étude, plusieurs montrent des patterns particulièrement informatifs :

Espèces bénéficiant fortement du LiDAR :

- **Grive draine** : gain modéré, variables de diversité dominantes
- **Sittelle torchepot** : faible gain, mais sélection cohérente avec l'écologie
- **Geai des chênes** : profil généraliste confirmé par les résultats

Ces observations soulignent la nécessité d'approches de conservation différenciées selon les exigences écologiques spécifiques plutôt que des stratégies uniformes pour tous les spécialistes forestiers.

L'ensemble de ces résultats révèle que la modélisation d'habitat aviaire par variables LiDAR est plus nuancée et complexe que les hypothèses théoriques ne le suggéraient initialement, nécessitant une approche pragmatique et adaptée aux spécificités de chaque espèce cible.

6. DISCUSSION

6.1 Interprétation écologique des résultats

Les résultats obtenus révèlent une réalité plus complexe que ne le suggéraient les hypothèses initiales. Si 57% des espèces bénéficient effectivement de l'intégration des variables LiDAR, la proportion importante d'espèces non-bénéficiaires (43%) interroge sur les mécanismes écologiques sous-jacents.

Il convient toutefois d'interpréter ces valeurs de gain avec parcimonie, compte tenu des biais méthodologiques identifiés dans les tests statistiques (voir annexe E). L'attention devrait plutôt être portée sur l'identification des variables LiDAR sélectionnées spécifiquement pour chaque espèce, qui révèle des patterns écologiques plus robustes que les seules comparaisons d'AUC. Toutefois, l'analyse des espèces "gagnantes" révèle des patterns instructifs. Le Coucou gris, avec le gain le plus substantiel (+0.0556), illustre paradoxalement que les espèces à écologie flexible peuvent être fortement dépendantes de la structure 3D. En tant que parasite de couvée, le Coucou gris doit identifier les nids de ses hôtes dans des configurations structurelles variées, expliquant potentiellement sa sensibilité aux métriques d'hétérogénéité verticale (Whelan, 2000). La Fauvette des jardins (+0.0382) et le Tarier pâtre (+0.0328), espèces de milieux semi-ouverts, confirment l'importance des variables d'ouverture de canopée pour les espèces de lisière, cohérent avec les observations de Tews et al. (2004) sur l'importance des structures clés du paysage.

À l'inverse, les résultats décevants pour des spécialistes emblématiques questionnent nos présupposés. Le Pic noir, malgré son statut de spécialiste des vieilles forêts et sa dépendance aux gros bois morts (Lammertink, 2007), ne montre aucun gain significatif (-0.007). Cette absence d'amélioration pourrait s'expliquer par plusieurs facteurs : d'une part, les variables LiDAR actuelles caractérisent mal la disponibilité en bois mort, ressource critique pour cette espèce ; d'autre part, la rareté du Pic noir (6,1% de prévalence) limite intrinsèquement la

robustesse statistique du modèle. Ces observations suggèrent que la spécialisation forestière ne se traduit pas automatiquement par une dépendance aux variables structurelles 3D

Remise en question du paradigme spécialiste/généraliste

L'hypothèse H2, prédisant une sensibilité accrue des spécialistes aux variables LiDAR, n'est que partiellement validée. Cette observation soulève des questions fondamentales sur la nature même de la spécialisation écologique. Les résultats suggèrent que la spécialisation d'habitat ne coïncide pas nécessairement avec la dépendance à la complexité structurelle 3D.

Plusieurs mécanismes peuvent expliquer cette discordance. Premièrement, certains spécialistes peuvent être davantage contraints par des facteurs non captés par le LiDAR : composition chimique de l'écorce (Stański et al., 2021), phénologie des ressources alimentaires, ou interactions biotiques. Deuxièmement, la spécialisation peut s'exprimer à des échelles spatiales différentes de celles caractérisées par les métriques LiDAR. Les données actuelles peuvent manquer des éléments de microhabitat critiques pour certaines espèces spécialisées.

Cette remise en perspective invite à redéfinir les critères de spécialisation structurelle au-delà de la dichotomie traditionnelle spécialiste/généraliste. Une approche multidimensionnelle, intégrant la spécialisation trophique, structurelle et spatiale, permettrait une compréhension plus fine des besoins écologiques aviaires.

Prédominance des variables "simples" : la robustesse écologique prime

L'invalidation de l'hypothèse H3 constitue l'un des résultats les plus instructifs de cette étude. La prédominance de CHM_Height_StdDev_m (écart-type des hauteurs) dans 67% des résultats, au détriment d'indices de complexité sophistiqués comme le Rumples Index ou la diversité de hauteur du feuillage, révèle que les oiseaux répondent davantage à des caractéristiques structurelles simples et robustes.

Cette observation trouve plusieurs explications écologiques. L'écart-type des hauteurs capture efficacement l'hétérogénéité verticale fondamentale qui détermine la diversification des niches aviaires (MacArthur & MacArthur, 1961), sans la complexité mathématique susceptible d'introduire du bruit dans le modèle. Cette métrique reflète directement les gradients de ressources et les opportunités de nidification stratifiées, concepts centraux en écologie aviaire (Holmes & Recher, 1986).

De plus, les variables simples présentent une robustesse statistique supérieure, particulièrement cruciale pour les espèces rares où la variabilité d'échantillonnage est élevée. Les indices de complexité, bien qu'écologiquement séduisants, peuvent être sensibles aux artefacts de mesure et aux conditions d'acquisition LiDAR, limitant leur applicabilité opérationnelle.

6.2 Comparaison avec la littérature internationale

Positionnement par rapport aux études similaires

Les gains d'AUC observés (médiane +0.016) s'inscrivent dans la gamme basse des améliorations rapportées dans la littérature internationale. Goetz et al. (2007) rapportent des améliorations plus substantielles (+0.05 à +0.15) pour la prédiction de richesse aviaire en forêt tempérée nord-américaine, tandis que Zellweger et al. (2013) observent des gains modérés (+0.02 à +0.08) en contexte alpin européen.

Cette variabilité s'explique par plusieurs facteurs contextuels. Premièrement, les écosystèmes forestiers wallons, fortement anthropisés et fragmentés, présentent une hétérogénéité structurelle potentiellement moins prononcée que les forêts anciennes nord-américaines. Deuxièmement, la densité élevée de points d'écoute en milieux de lisière dans le protocole SOCWAL peut atténuer l'importance relative des variables de structure de cœur de forêt. Enfin, la variabilité inter-spécifique importante observée dans cette étude contraste avec les approches de richesse ou diversité totale privilégiées dans d'autres travaux.

Variables LiDAR importantes : convergence et spécificités régionales

La prédominance de CHM_Height_StdDev_m trouve des échos dans plusieurs études européennes. Bergen et al. (2009) identifient les métriques de variabilité des hauteurs comme prédicteurs robustes de la diversité aviaire, cohérent avec nos observations. Cependant, l'importance de Forest_Edge_Density_per_m2 (59% des espèces) constitue potentiellement une spécificité du jeu de données.

Cette prédominance des variables de lisière pourrait refléter la forte fragmentation du paysage forestier wallon (Alderweireld et al., 2015) et la concentration des points d'écoute SOCWAL en zones accessibles, souvent périphériques. Cette observation souligne l'importance du contexte

paysager dans l'interprétation des résultats et limite potentiellement leur transférabilité à des contextes forestiers moins fragmentés.

6.3 Limites méthodologiques et autocritique

Limitations structurelles des données d'entrée

L'organisation des données SOCWAL en chaînes de points d'écoute constitue la limitation la plus contraignante de cette étude. Cette structure génère une autocorrélation spatiale extrême pour certaines espèces, dépassant les capacités de gestion par la validation croisée spatiale. Sept espèces présentent des portées d'autocorrélation supérieures à 30 km, rendant l'indépendance statistique difficile à garantir.

Cette limitation n'est pas anecdotique : elle affecte directement la fiabilité des estimations de performance et peut conduire à un optimisme artificiel pour les espèces à distribution très agrégée. Les 11% d'espèces à qualité de folds problématique témoignent de ces contraintes structurelles, particulièrement critiques pour les espèces de plus grand intérêt conservatoire.

Le biais de détection différentielle inhérent aux points d'écoute de 5 minutes constitue une seconde limitation majeure. Cette méthodologie privilégie intrinsèquement les passereaux vocalisant activement au détriment des rapaces, espèces nocturnes et espèces discrètes. Cette sélectivité limite de facto l'étude à un sous-ensemble taxonomique de l'avifaune wallonne, excluant des guildes écologiques potentiellement sensibles aux variables structurelles 3D.

Choix méthodologiques : compromis et alternatives

L'approche adaptative VSURF, bien que innovante, présente des compromis inhérents entre robustesse et sensibilité. La configuration ultra-conservatrice pour les espèces très rares (seuil de consensus 80%) peut conduire à une sous-sélection par excès de parcimonie, masquant potentiellement des relations écologiques authentiques mais faibles. Inversement, la configuration relaxée pour les espèces communes risque la sur-sélection, particulièrement problématique en présence de variables corrélées.

Le choix des seuils de prévalence (15% et 5%) constitue un autre point de débat méthodologique. Ces seuils, bien qu'appuyés par la littérature (Japkowicz & Stephen, 2002 ;

Breiner et al., 2015), demeurent arbitraires et influencent directement les résultats. Une analyse de sensibilité sur ces seuils aurait renforcé la robustesse des conclusions.

Perfectibilité du modèle développé

Il convient de reconnaître avec transparence que le modèle présenté n'est pas parfait. Les contraintes temporelles de développement ont limité l'optimisation fine de certains paramètres, et plusieurs biais méthodologiques persistent malgré les précautions prises. Cependant, cette étude constitue une étape importante vers un modèle encore plus performant.

Le pipeline adaptatif développé pose les bases méthodologiques pour des améliorations futures : intégration de données multitemporelles, raffinement des métriques LiDAR écologiquement pertinentes, couplage avec d'autres sources de télédétection. La philosophie adaptative selon la prévalence des espèces représente un progrès conceptuel, ouvrant la voie à des approches encore plus sophistiquées.

Malgré les limitations observées, les résultats restent remarquables pour une modélisation automatisée de données écologiques complexes. La cohérence entre les résultats du modèle et l'écologie connue des espèces renforce la validité de l'approche, les cas problématiques révélant davantage des pistes d'optimisation que des défauts fondamentaux de la méthode.

6.4 Implications pour la gestion forestière

Recommandations opérationnelles prioritaires

Les résultats fournissent des orientations concrètes pour la gestion forestière wallonne. La prédominance de l'hétérogénéité des hauteurs comme déterminant structural universel se traduit par une recommandation sylvicole claire : éviter les peuplements mono-strates au profit de structures verticalement diversifiées.

Cette recommandation implique concrètement le maintien d'arbres de diverses classes d'âge lors des interventions sylvicoles, la création de trouées de régénération de tailles variées, et la conservation d'arbres dominants isolés dans les zones d'éclaircie. L'importance de `Forest_Edge_Density_per_m2` pour 59% des espèces justifie une gestion différenciée des

lisières, avec création de gradients progressifs plutôt que de transitions abruptes forêt-milieu ouvert.

Ces recommandations trouvent une traduction opérationnelle directe dans les pratiques sylvicoles : adoption de régimes de futaie irrégulière ou jardinée, modulation de l'intensité des éclaircies selon les objectifs de biodiversité, intégration d'îlots de sénescence pour maintenir l'hétérogénéité structurelle.

Contexte réglementaire européen et indicateurs de suivi

Dans le contexte du Règlement européen sur la restauration de la nature, qui oblige les pays de l'UE à renverser d'ici 2030 la tendance au déclin des indicateurs des oiseaux forestiers, cette étude apporte une modeste contribution. L'identification de CHM_Height_StdDev_m comme métrique structurelle universelle offre un indicateur LiDAR opérationnel pour le suivi de la qualité des habitats forestiers.

Cette approche permettrait de compléter les indicateurs démographiques classiques par des métriques d'habitat, facilitant ainsi l'évaluation proactive des mesures de restauration pour chaque espèce. L'intégration de ces variables dans les protocoles de suivi constituerait une avancée pour l'évaluation de l'efficacité des politiques de conservation forestière, entre autres, et permettrait une meilleure compréhension des habitats des différentes espèces.

Stratégies différenciées selon les profils écologiques

Les résultats justifient l'abandon d'approches uniformes au profit de stratégies de conservation différenciées. Les espèces "LiDAR-dépendantes" nécessitent une attention particulière à la structure 3D des habitats, tandis que les espèces "LiDAR-indépendantes" requièrent probablement des approches centrées sur d'autres facteurs (composition floristique, ressources trophiques, utilisation du sol,...).

Cette différenciation implique le développement d'outils d'aide à la décision permettant aux gestionnaires d'adapter leurs pratiques selon les espèces cibles prioritaires.

6.5 Perspectives méthodologiques et enjeux de recherche future

Améliorations du pipeline développé

Plusieurs axes d'amélioration émergent pour perfectionner le pipeline développé.

L'intégration de données LiDAR multitemporelles permettrait de caractériser la dynamique structurelle et d'améliorer la robustesse des prédictions. Le couplage avec des données météorologiques et phénologiques affinerait la compréhension des relations temporelles habitat-espèce.

L'optimisation des métriques LiDAR constitue un axe prioritaire. Le développement d'indices spécifiquement conçus pour caractériser les ressources aviaires, telles que la disponibilité des sites de nidification ou la distribution des ressources trophiques, pourrait améliorer significativement les performances prédictives du modèle. Il serait également pertinent d'identifier une variable adaptée pour quantifier la présence de bois mort, ressource clé pour de nombreuses espèces forestières.

Une perspective particulièrement prometteuse serait l'intégration de l'approche ITC (Individual Tree Crown) en complément de l'approche ABA utilisée dans ce travail. Contrairement aux métriques raster agrégées employées ici, l'approche ITC permet la détection et la cartographie individuelle des houppiers, ouvrant la voie à une caractérisation plus fine de la structure forestière. Cette méthode permettrait notamment d'identifier et de cartographier les gros arbres, ressources critiques pour de nombreuses espèces cavernicoles comme le Pic noir ou la Sittelle torchepot. L'approche ITC fournirait des métriques complémentaires telles que la densité d'arbres de gros diamètre par hectare, la distribution spatiale des arbres dominants, ou encore la caractérisation précise de l'architecture des houppiers, variables potentiellement déterminantes pour les espèces spécialistes mais non captées par les métriques raster actuelles.

Innovations technologiques et méthodologiques

L'émergence des données LiDAR à très haute densité (>20 points/m²) et des acquisitions multisaisonnières ouvre de nouvelles perspectives pour la caractérisation fine des habitats aviaires. L'intégration de l'intelligence artificielle et du deep learning pour l'analyse automatisée de nuages de points 3D pourrait révolutionner l'extraction de métriques écologiquement pertinentes.

Le couplage LiDAR-données hyperspectrales permettrait d'intégrer simultanément structure et composition chimique de la végétation, offrant une caractérisation holistique des habitats. Ces innovations technologiques nécessiteront cependant des adaptations méthodologiques substantielles.

Questions scientifiques émergentes

Cette étude ouvre plusieurs questions de recherche fondamentales. Les mécanismes fins de perception et d'utilisation de la structure 3D par les oiseaux demeurent largement méconnus. L'intégration des effets du changement climatique sur les relations structure-avifaune constitue un enjeu majeur pour la gestion adaptative des forêts.

6.6 Portée et limites de transférabilité

Validité géographique et temporelle

La transférabilité géographique des résultats nécessite une évaluation prudente. Le contexte forestier wallon, caractérisé par une forte fragmentation et une gestion intensive, peut limiter la généralisation à d'autres régions européennes. La validation sur des écosystèmes forestiers contrastés constituerait une étape essentielle pour évaluer la robustesse des patterns identifiés.

La validité temporelle pose également question dans un contexte de changement climatique rapide. Les relations structure-habitat identifiées pour la période 2019-2022 peuvent évoluer avec les modifications de phénologie et de distribution des espèces. Le suivi à long terme de ces relations constitue un enjeu majeur pour la gestion adaptative.

Implications pour la recherche européenne

Cette étude vise à contribuer à l'effort européen de standardisation des méthodes d'évaluation de la biodiversité forestière. L'harmonisation des protocoles LiDAR et le développement d'indicateurs communs faciliteraient les comparaisons transfrontalières et l'évaluation des politiques européennes de conservation.

L'intégration dans les réseaux de surveillance existants permettrait de valoriser ces approches à l'échelle continentale et de contribuer aux objectifs de la Stratégie européenne pour la biodiversité 2030.

Cette discussion révèle que l'apport des variables LiDAR à la modélisation d'habitat aviaire, bien que significatif pour une majorité d'espèces, s'avère plus nuancé et contexte-dépendant que ne le suggéraient les hypothèses initiales. Ces résultats ouvrent néanmoins des perspectives prometteuses pour une gestion forestière plus fine et écologiquement informée, à condition d'adapter les approches aux spécificités de chaque espèce et contexte écologique.

7. Contribution personnelle

Afin de mener à bien ce travail de recherche, l'étudiante a été amenée à réaliser toute une série d'activités enrichissantes et méthodologiquement exigeantes. Elle a procédé à un état des lieux approfondi des connaissances sur la modélisation d'habitat aviaire et l'utilisation du LiDAR en écologie via des recherches bibliographiques poussées, explorant tant la littérature francophone qu'internationale pour maîtriser les enjeux théoriques et méthodologiques complexes de ce domaine en pleine évolution.

L'étudiante a développé de manière autonome des compétences informatiques et statistiques avancées, via l'utilisation du logiciel QGIS et du langage R avec RStudio pour l'ensemble du pipeline de modélisation.

L'étudiante a harmonisé des données multi-sources hétérogènes couvrant un gradient environnemental quasi-complet : données LiDAR haute résolution, observations aviaires SOCWAL, composition forestière ForestImator, et données d'occupation du sol WALOUS. Cette intégration a permis de dépasser le cadre uniquement forestier puisque l'approche multi-habitat utilisée a couvert les espèces urbaines et agricoles en plus des forestières.

L'étudiante a cerné les attentes des personnes pouvant bénéficier du travail et a adapté la mise en forme des résultats pour convenir aux principaux intéressés : d'un côté les résultats académiques rigoureux, de l'autre les résultats utilisables directement sur le terrain par les gestionnaires forestiers et les acteurs de la conservation.

Enfin, elle a fait preuve de réflexion et d'esprit critique dans l'interprétation des résultats, reconnaissant avec transparence les limitations méthodologiques tout en valorisant les apports concrets pour la gestion forestière.

8. CONCLUSION générale

Le LiDAR n'est pas la panacée universelle attendue, mais constitue un outil précieux pour une gestion forestière écologiquement informée, à condition d'adapter son usage aux spécificités de chaque espèce.

Cette étude démontre que la révolution technologique LiDAR, pour être pleinement opérationnelle, doit s'accompagner d'une révolution conceptuelle : abandonner les approches uniformes au profit de stratégies différenciées selon les profils écologiques. L'enjeu n'est plus de savoir si le LiDAR améliore la modélisation d'habitat, mais pour quelles espèces, dans quels contextes, et selon quelles modalités.

En révélant ces nuances, cette recherche contribue modestement mais concrètement à une gestion forestière plus fine, mieux adaptée aux exigences de conservation de notre avifaune dans un contexte de changements environnementaux accélérés. La forêt vue du ciel nous enseigne finalement l'humilité : chaque espèce a sa propre manière d'habiter l'espace tridimensionnel, et c'est en respectant cette diversité d'exigences que nous pourrions concilier gestion forestière et conservation de la biodiversité.

RÉFÉRENCES BIBLIOGRAPHIQUES

Alderweireld M., Burnay F., Pitchugin P. & Lecomte H. (2015) Inventaire forestier wallon - Résultats 1994-2012. SPW, DGO3, Direction des Ressources Forestières, Jambes : 236 p.

Araújo, M. B., & Peterson, A. T. (2012). Uses and misuses of bioclimatic envelope modeling. *Ecology*, 93(7), 1527-1539. <https://doi.org/10.1890/11-1930.1>

Austin, M. (2007). Species distribution models and ecological theory: a critical assessment and some possible new approaches. *Ecological Modelling*, 200(1-2),.

Barbet-Massin, M., Jiguet, F., Albert, C. H., & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution*, 3(2), 327-338.

Bergen, K. M., Goetz, S. J., Dubayah, R. O., Henebry, G. M., Hunsaker, C. T., Imhoff, M. L., Nelson, R. F., Parker, G. G., Radeloff, V. C., & Steinberg, D. H. (2009). Remote sensing of vegetation 3-D structure for biodiversity and habitat: Review and implications for lidar and radar spaceborne missions. *Journal of Geophysical Research: Biogeosciences*, 114(G2).

Bolyn, C., Lejeune, P., Michez, A., & Latte, N. (2022). Mapping tree species proportions from satellite imagery using spectral-spatial deep learning. *Remote Sensing of Environment*, 280, 113205. <https://doi.org/10.1016/j.rse.2022.113205>

Both, C., Bouwhuis, S., Lessells, C. M., & Visser, M. E. (2006). Climate change and population declines in a long-distance migratory bird. *Nature*, 441(7089), 81-83. <https://doi.org/10.1038/nature04539>

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>

Breining, D. R., Burgman, M. A., & Stoms, D. M. (1998). Analyzing habitat relationships of endangered birds using GIS and statistical models. *Journal of Wildlife Management*, 62(3), 995-1006.

Breiner, F. T., Guisan, A., Bergamini, A., & Nobis, M. P. (2015). Overcoming limitations of modelling rare species by using ensembles of small models. *Methods in Ecology and Evolution*, 6(10), 1210-1218. <https://doi.org/10.1111/2041-210X.12403>

Burgman, M. A., Lindenmayer, D. B., & Elith, J. (2005). Managing landscapes for conservation under uncertainty. *Ecology*, 86(8), 2007-2017.

Chen, C., Liaw, A., & Breiman, L. (2004). Using random forest to learn imbalanced data. *University of California, Berkeley*, 110, 1-12.

Clavel, J., Julliard, R., & Devictor, V. (2011). Worldwide decline of specialist species: toward a global functional homogenization? *Frontiers in Ecology and the Environment*, 9(4), 222-228.

Cockle, K. L., Capuzzi, G., Bodrati, A., Clay, R. P., Del Castillo, H., Velazquez, M., Politi, N., & Di Sallo, F. (2015). Cavity characteristics, but not habitat, influence nest survival of cavity-nesting birds along a gradient of human impact in South American temperate forests. *Biological Conservation*, 227, 84-93.

<https://doi.org/10.1016/j.biocon.2015.01.026>

Cook, C. N., Mascia, M. B., Schwartz, M. W., Possingham, H. P., & Fuller, R. A. (2013). Achieving conservation science that bridges the knowledge-action boundary. *Conservation Biology*, 27(4), 669-678.

<https://doi.org/10.1111/cobi.12050>

Cressie, N., & Wikle, C. K. (2015). *Statistics for Spatio-Temporal Data*. Wiley. ISBN: 978-0-471-69274-4

Davies, A. B., & Asner, G. P. (2014). Advances in animal ecology from 3D-LiDAR ecosystem mapping. *Trends in Ecology & Evolution*, 29(12), 681-691. <https://doi.org/10.1016/j.tree.2014.10.005>

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 44(3), 837-845.

Derouaux, A. (2021). *Suivi des oiseaux communs en Wallonie 1990-2020. Rapport aux observateurs*. Aves/Natagora, 64 pp.

Derouaux, A., & Paquet, J.-Y. (2018). L'évolution préoccupante des populations d'oiseaux nicheurs en Wallonie : 28 ans de surveillance de l'avifaune commune. *Aves*, 55(1), 1-31.

Devictor, V., Julliard, R., & Jiguet, F. (2008). Distribution of specialist and generalist species along spatial gradients of habitat disturbance and fragmentation. *Oikos*, 117(4), 507-514.

Dormann, C. F., McPherson, J. M., Araújo, M. B., Bivand, R., Bolliger, J., Carl, G., Davies, R. G., Hirzel, A., Jetz, W., Kissling, W. D., Kühn, I., Ohlemüller, R., Peres-Neto, P. R., Reineking, B., Schröder, B., Schurr, F. M., & Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, 30(5), 609-628.

<https://doi.org/10.1111/j.2007.0906-7590.05171.x>

Elith, J. (2000). Quantitative methods for modeling species habitat: comparative performance and an application to Australian plants. Dans S. Ferson & M. Burgman (Éds.), *Quantitative methods for conservation biology* (pp. 39-58). Springer.

Faaborg, J., Holmes, R. T., Anders, A. D., Bildstein, K. L., Dugger, K. M., et al. (2010). Recent advances in understanding migration systems of New World land birds. *Ecological Monographs*, 80, 3-48.

<https://doi.org/10.1890/09-0395.1>

Fahrig, L. (2003). Effects of habitat fragmentation on biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, 34(1), 487-515.

Ferrier, S., & Watson, G. (1997). An evaluation of the effectiveness of environmental surrogates and modelling techniques for predicting the distribution of biological diversity. *Environment Australia*, Canberra.

Ferrier, S., Watson, G., Pearce, J., & Drielsma, M. (2002). Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Species-level modelling. *Biodiversity & Conservation*, 11(12), 2275-2307.

Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24(1), 38-49.

Fortin, M. J., & Dale, M. (2005). *Spatial analysis: a guide for ecologists*. Cambridge University Press.

Fowler, M. (2010). *Domain-specific languages*. Addison-Wesley Professional.

Genuer, R., Poggi, J. M., & Tuleau-Malot, C. (2015). VSURF: an R package for variable selection using random forests. *The R Journal*, 7(2), 19-33. DOI:10.32614/RJ-2015-018

Goetz, S., Steinberg, D., Dubayah, R., & Blair, B. (2007). Laser remote sensing of canopy habitat heterogeneity as a predictor of bird species richness in an eastern temperate forest, USA. *Remote Sensing of Environment*, 108(3), 254-263.

Gumede, S. T., Ehlers Smith, Y. C., Ehlers Smith, D. A., Downs, C. T., & Nsor, C. A. (2022). Avian species functional diversity and habitat use: The role of forest structural attributes and tree diversity in the Midlands Mistbelt Forests of KwaZulu-Natal, South Africa. *Ecology and Evolution*, 13(9), e10439.

<https://doi.org/10.1002/ece3.10439>

Gunnarsson, B., Nittérus, K., & Wirdeñäs, P. (2011). Management effect on bird and arthropod interaction in suburban woodlands. *BMC Ecology*, 11(1), 1-9.

Halme, E., & Niemelä, J. (1993). Carabid beetles in fragments of coniferous forest. *Annales Zoologici Fennici*, 30(1), 17-30.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer.

- Hinsley, S. A., Hill, R. A., Bellamy, P. E., Harrison, N. M., Speakman, J. R., Wilson, A. K., & Ferns, P. N. (2002). A bird's eye view: using remote sensing to gauge the importance of urban green space for birds. *Ibis*, 144(4), 639-648.
- Holmes, R. T., & Recher, H. F. (1986). Determinants of guild structure in forest bird communities: an intercontinental comparison. *The Condor*, 88(4), 427-439.
- Holmes, R. T., & Robinson, S. K. (1981). Tree species preferences of foraging insectivorous birds in a northern hardwoods forest. *Oecologia*, 48(1), 31-35.
- Hyypä, J., Hyypä, H., Leckie, D., Gougeon, F., Yu, X., & Maltamo, M. (2008). Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. *International Journal of Remote Sensing*, 29(5), 1339-1366. <https://doi.org/10.1080/01431160701736489>
- Japkowicz, N., & Stephen, S. (2002). The class imbalance problem: A systematic study. *Intelligent Data Analysis*, 6(5), 429-449.
- Jiménez-Valverde, A., Peterson, A. T., Soberón, J., Overton, J. M., Aragón, P., & Lobo, J. M. (2011). Use of niche models in invasive species risk assessments. *Biological Invasions*, 13(12), 2785-2797. <https://doi.org/10.1007/s10530-011-9963-4>
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, 2, 1137-1143.
- Korhonen, L., Korpela, I., Heiskanen, J., & Maltamo, M. (2011). Airborne discrete-return LIDAR data in the estimation of vertical canopy cover, angular canopy closure and leaf area index. *Remote Sensing of Environment*, 115(4), 1065-1078. <https://doi.org/10.1016/j.rse.2010.12.011>
- Lammertink, M. (2007). Foraging on arboreal ants by North American woodpeckers. *The Wilson Journal of Ornithology*, 119(3), 429-435.
- Latte, N., & Claessens, H. (2014). Impact of climate change on forest growth and wood production in southern Belgium. *Forest Ecology and Management*, 328, 364-373.
- Lefsky, M. A., Cohen, W. B., Acker, S. A., Parker, G. G., Spies, T. A., & Harding, D. (1999). Lidar remote sensing of the canopy structure and biophysical properties of Douglas-fir western hemlock forests. *Remote Sensing of Environment*, 70(3), 339-361.
- Lindenmayer, D. B., & Likens, G. E. (2010). *Effective Ecological Monitoring*. CSIRO Publishing and Earthscan, Melbourne and London.

Listopad, C. M., Masters, R. E., Drake, J., Weishampel, J., & Branquinho, C. (2015). Structural diversity indices based on airborne LiDAR as ecological indicators for managing forest ecosystems. *Ecological Indicators*, 57, 171-179. <https://doi.org/10.1016/j.ecolind.2015.04.017>

Lomba, A., Pellissier, L., Randin, C., Vicente, J., Moreira, F., Honrado, J., & Guisan, A. (2010). Overcoming the rare species modelling paradox: a novel hierarchical framework applied to an Iberian endemic plant. *Biological Conservation*, 143(11), 2647-2657.

MacArthur, R. H., & MacArthur, J. W. (1961). On bird species diversity. *Ecology*, 42(3), 594-598. <https://doi.org/10.2307/1932254>

Magnussen, S., & Boudewyn, P. (1998). Derivations of stand heights from airborne laser scanner data with canopy-based quantile estimators. *Canadian Journal of Forest Research*, 28(7), 1016-1031.

Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, 58(8), 1246-1266.

McClure, C. J. W., Rolek, B. W., & Hill, G. E. (2012). Predicting Occupancy of Wintering Migratory Birds: is Microhabitat Information Necessary? *The Condor*, 114(3), 482-490.

Meinshausen, N., & Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society: Series B*, 72(4), 417-473.

Michez, A., Morelle, K., Lehaire, F., Wieland, R., Lejeune, P., Vermeulen, C., & Latte, N. (2022). Mapping forest species composition from Sentinel-2 time series and airborne LiDAR acquisitions in a temperate broadleaf forest. *Remote Sensing of Environment*, 280, 113145. <https://doi.org/10.1016/j.rse.2022.113145>

Morsdorf, F., Kötz, B., Meier, E., Itten, K. I., & Allgöwer, B. (2006). Estimation of LAI and fractional cover from small footprint airborne laser scanning data based on gap fraction. *Remote Sensing of Environment*, 104(1), 50-61.

Næsset, E. (2002). Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment*, 80(1), 88-99.

Newton, I. (2008). *The Migration Ecology of Birds*. Academic Press, Waltham, Massachusetts. ISBN: 978-0-12-517367-4

Oshiro, T. M., Perez, P. S., & Baranauskas, J. A. (2012). How many trees in a random forest? Dans *International workshop on machine learning and data mining in pattern recognition* (pp. 154-168). Springer.

Parker, G. G., & Russ, M. E. (2004). The canopy surface and stand development: assessing forest canopy structure and complexity with near-surface altimetry. *Forest Ecology and Management*, 189(1-3), 307-315.

- Reading, R. P., Clark, T. W., & Griffith, B. (1996). The influence of valuational and organizational considerations on the success of rare species translocations. *Biological Conservation*, 79(1), 25-31.
- Reichman, O. J., Jones, M. B., & Schildhauer, M. P. (2011). Challenges and opportunities of open data in ecology. *Science*, 331(6018), 703-705. <https://doi.org/10.1126/science.1197962>
- Roberts, D. R., et al. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40, 913-929. <https://doi.org/10.1111/ecog.02881>
- Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One*, 10(3), e0118432.
- Schratz, P., Muenchow, J., Iturrutxa, E., Richter, J., & Brenning, A. (2019). Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. *Ecological Modelling*, 406, 109-120. <https://doi.org/10.1016/j.ecolmodel.2019.06.002>
- Shaw, D. C., Bible, K., Freund, J. A., Yost, A. C., Wayman, R. B., Ringo, C., & Murray, K. (2024). Forest structural heterogeneity positively affects bird richness and acoustic diversity in a temperate, central European forest. *Frontiers in Ecology and Evolution*, 12, 1387879. <https://doi.org/10.3389/fevo.2024.1387879>
- Simonson, W. D., Allen, H. D., & Coomes, D. A. (2014). Applications of airborne lidar for the assessment of animal species diversity. *Methods in Ecology and Evolution*, 5(8), 719-729. <https://doi.org/10.1111/2041-210X.12219>
- Stański, T., Stańska, M., Goławski, A., & Czeszczewik, D. (2021). Foraging site selection of the middle spotted woodpecker (*Leiopicus medius* L.) in primeval oak-lime-hornbeam forest of the Białowieża National Park: comparison of breeding and non-breeding seasons. *Forests*, 12(7), 837. <https://doi.org/10.3390/f12070837>
- Stenseth, N. C., Mysterud, A., Ottersen, G., Hurrell, J. W., Chan, K. S., & Lima, M. (2002). Ecological effects of climate fluctuations. *Science*, 297(5585), 1292-1296. <https://doi.org/10.1126/science.1071281>
- Strobl, C., Boulesteix, A. L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinformatics*, 8(1), 25.
- Strobl, C., Boulesteix, A. L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, 9(1), 307.
- Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240(4857), 1285-1293.

- Telford, R. J., & Birks, H. J. B. (2005). The secret assumption of transfer functions: problems with spatial autocorrelation in evaluating model performance. *Quaternary Science Reviews*, 24(20-21), 2173-2179.
- Tews, J., Brose, U., Grimm, V., Tielbörger, K., Wichmann, M. C., Schwager, M., & Jeltsch, F. (2004). Animal species diversity driven by habitat heterogeneity/diversity: the importance of keystone structures. *Journal of Biogeography*, 31(1), 79-92.
- Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillerá-Arroita, G. (2019). blockCV: an R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution*, 10(2), 225-232.
<https://doi.org/10.1111/2041-210X.13107>
- Van Horne, B., & Wiens, J. A. (1991). Forest bird habitat suitability models and the development of general habitat models. *US Fish and Wildlife Service, Fish and Wildlife Research* 8.
- Vierling, K. T., Vierling, L. A., Gould, W. A., Martinuzzi, S., & Clawges, R. M. (2008). Lidar: shedding new light on habitat characterization and modeling. *Frontiers in Ecology and the Environment*, 6(2), 90-98.
- Webster, R., & Oliver, M. A. (2007). *Geostatistics for environmental scientists* (2nd ed.). John Wiley & Sons.
- Wehr, A., & Lohr, U. (1999). Airborne laser scanning—an introduction and overview. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54(2-3), 68-82.
- Whelan, C. J. (2000). *Bird community composition and habitat selection in forest ecosystems*. Academic Press.
- Willson, M. F. (1974). Avian community organization and habitat structure. *Ecology*, 55(5), 1017-1029.
- Wintle, B. A., Elith, J., & Potts, J. M. (2005). Fauna habitat modelling and mapping: A review and case study in the Lower Hunter Central Coast region of NSW. *Austral Ecology*, 30(7), 719-738.
- Yates, K. L., Bouchet, P. J., Caley, M. J., Mengersen, K., Randin, C. F., Parnell, S., ... & Sequeira, A. M. M. (2018). Outstanding challenges in the transferability of ecological models. *Trends in Ecology & Evolution*, 33(10), 790-802. <https://doi.org/10.1016/j.tree.2018.08.001>
- Zaniewski, A. E., Lehmann, A., & Overton, J. M. (2002). Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modelling*, 157(2-3), 261-280.
- Zellweger, F., Braunisch, V., Baltensweiler, A., & Bollmann, K. (2013). Remotely sensed forest structural complexity predicts multi-species occurrence at the landscape scale. *Forest Ecology and Management*, 307, 303-312. <https://doi.org/10.1016/j.foreco.2013.07.023>

Sites Internet et Ressources Numériques

ForestImator Platform. (2023). Forest composition mapping for Wallonia. Université de Liège - Gembloux Agro-Bio Tech. Consulté le 27/04/2025 sur <https://forestimator.gembloux.ulg.ac.be/>

Géoportail de la Wallonie. Service Public de Wallonie. Consulté le 28/05/2025 sur <https://geodata.wallonie.be/>

Service public de Wallonie (SPW), Université Libre de Bruxelles (ULB), Université catholique de Louvain (UCL), Institut Scientifique de Service Public (ISSeP). (2020). Utilisation du Sol en Wallonie - WALOUS 2018. Consulté le 28/05/2025 sur <https://geodata.wallonie.be/id/fafb9fc7-4e32-4a40-8163-c64966cb573c>

Service Public de Wallonie (SPW). (2023). Géoportail de la Wallonie - Données LiDAR 2021-2022. SPW Digital - Département Données transversales - Production géomatique et traitement de données. <https://geoportail.wallonie.be/lidar>.

Contact technique : lidar@spw.wallonie.be

Outils d'Intelligence Artificielle

Les modèles d'intelligence artificielle suivants ont été consultés pour l'assistance à la rédaction, la vérification de syntaxe et l'organisation du contenu :

Claude (version 3.5 Sonnet) : Assistance à la structuration méthodologique et mise en forme du document. Consulté via <https://claude.ai/>

OpenAI ChatGPT (version GPT-4) : Support pour le codage, la gestion des erreurs de code, l'optimisation de la présentation des résultats et l'aide à la correction orthographique, l'auteur de ce mémoire étant dysorthographique. Consulté via <https://chat.openai.com/>.

Déclaration de transparence : Conformément aux bonnes pratiques académiques, je déclare avoir utilisé ces outils d'IA générative comme aide à la rédaction, à l'organisation du contenu, à la mise en forme et à la correction orthographique, tout en conservant l'entière responsabilité ainsi que le crédit intellectuel du contenu scientifique, des choix méthodologiques, de l'analyse des données, de l'interprétation des résultats et des conclusions présentées dans ce mémoire.

Logiciels et Environnements de Développement

QGIS (version 3.34) : Système d'information géographique open source utilisé pour le traitement des données spatiales, la visualisation cartographique des données LiDAR et d'occupation du sol. QGIS Development Team (2024). Disponible sur <https://qgis.org/>

R (version 4.3.2) : Langage et environnement de programmation statistique utilisé pour l'ensemble des analyses de données, la modélisation Random Forest, l'implémentation de VSURF et les analyses statistiques. R Core Team (2024). R Foundation for Statistical Computing, Vienna, Austria. Disponible sur <https://www.R-project.org/>

RStudio (version 2023.12.1) : Environnement de développement intégré pour R utilisé pour l'ensemble du pipeline de modélisation, de la préparation des données à la génération des résultats et visualisations. RStudio Team (2024). Posit Software, PBC, Boston, MA. Disponible sur <https://www.rstudio.com/>

Packages R principaux utilisés :

- **randomForest** (v4.7-1) : Implémentation de l'algorithme Random Forest
- **VSURF** (v1.2.0) : Sélection de variables adaptative
- **blockCV** (v3.1-4) : Validation croisée spatiale
- **sf** (v1.0-14) : Manipulation des données spatiales vectorielles
- **gstat** : Analyse géostatistique et variogrammes
- **pROC** : Calcul des métriques de performance (AUC-ROC)
- **doParallel** : Parallélisation des calculs

ANNEXES

A.1 Comparaison avec la campagne précédente

Paramètre	Campagne 2013-2014	Campagne 2021-2022	Facteur d'amélioration
Densité de points	1,51 pulse/m ²	6,8 pulse/m ²	×4,5
Résolution spatiale	Standard	Haute résolution	Significative

Note importante : L'amélioration substantielle de la densité de points entre les deux campagnes (facteur 4,5) rend les données méthodologiquement incompatibles pour les analyses comparatives temporelles directes.

Annexe B : Spécifications techniques des métriques LiDAR

B.1 Paramètres généraux de traitement

Configuration du système

- **Traitement parallèle :** 6 workers (`plan(multisession, workers = 6)`)
- **Parcelles traitées :** 2469 parcelles circulaires de 300m de rayon
- **Normalisation :** Automatique des nuages de points LiDAR
- **Filtrage :** Hauteurs réalistes (0.5-60m pour végétation)
- **Gestion d'erreurs :** `tryCatch()` pour chaque bloc de calcul

Matrice de Corrélation - Toutes Variables

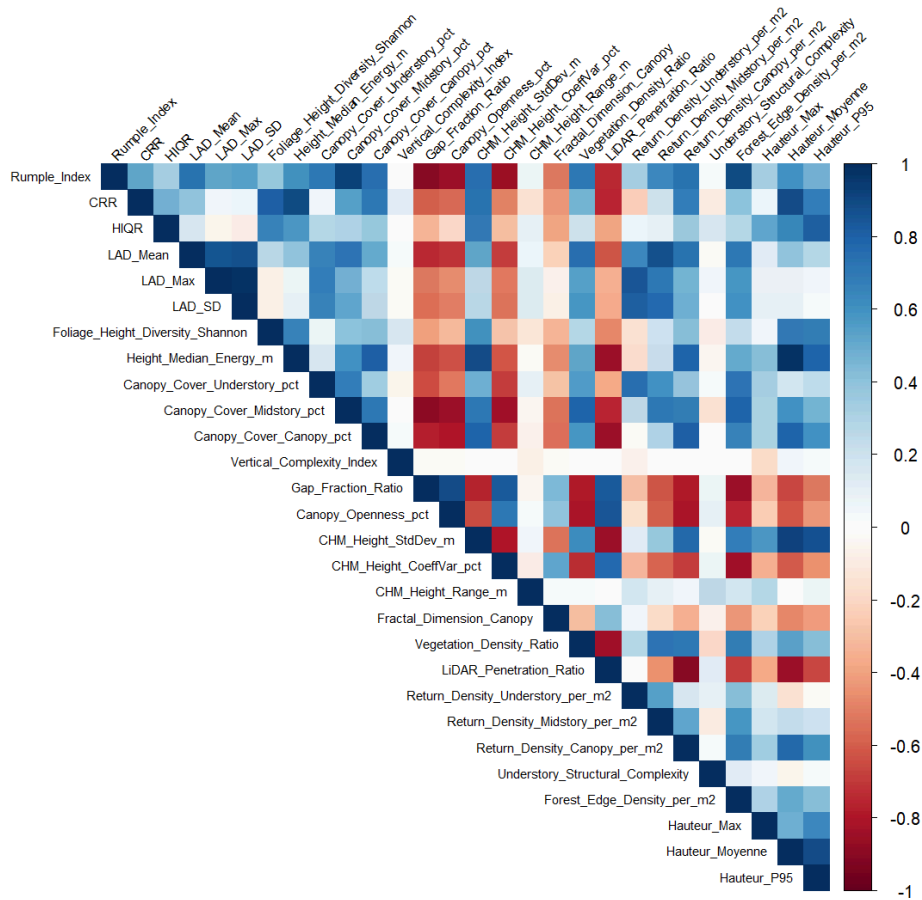


Figure 3 : Matrice de corrélation des variables LiDAR

B.2 Métriques de hauteur (n = 8)

B.2.1 Variables basées sur les points natifs

Hauteur_Max - Hauteur maximale

Résolution : Points natifs (6.8 pts/m²) Interprétation : Hauteur dominante de la canopée

Hauteur_Moyenne - Hauteur moyenne

Résolution : Points natifs *Interprétation* : Hauteur moyenne pondérée par la densité de végétation

Hauteur_P95 - 95e percentile des hauteurs

Résolution : Points natifs *Interprétation* : Hauteur dominante robuste aux valeurs aberrantes

Height_Median_Energy_m - Médiane énergétique (HOME)

Résolution : Points natifs *Interprétation* : Centre de masse vertical de la végétation

HIQR - Écart interquartile des hauteurs

Résolution : Points natifs *Interprétation* : Variabilité centrale des hauteurs

B.2.2 Variables basées sur le CHM (raster 1m)

CHM_Height_StdDev_m - Écart-type des hauteurs CHM

Résolution : Raster 1m × 1m *Interprétation* : Hétérogénéité structurelle horizontale

CHM_Height_CoeffVar_pct - Coefficient de variation

Résolution : Raster 1m × 1m *Interprétation* : Variabilité relative normalisée (plafonné à 500%)

CHM_Height_Range_m - Étendue des hauteurs

Résolution : Raster 1m × 1m *Interprétation* : Amplitude totale de variation des hauteurs

B.3 Métriques de densité (n = 12)

B.3.1 Densité foliaire (LAD)

LAD_Mean, LAD_Max, LAD_SD - Leaf Area Density

Résolution : 10 strates verticales équidistantes *Interprétation* : Distribution verticale de la biomasse foliaire

B.3.2 Densités par strates

Return_Density_Understory_per_m2 - Densité sous-bois

Strate : 0.5-5m | *Résolution* : Points natifs/surface parcelle

Return_Density_Midstory_per_m2 - Densité étage moyen

Strate : 5-15m | *Résolution* : Points natifs/surface parcelle

Return_Density_Canopy_per_m2 - Densité canopée

Strate : 15-60m | *Résolution* : Points natifs/surface parcelle

B.3.3 Ratios de densité

Vegetation_Density_Ratio - VDR

Interprétation : Ratio végétation/sol, indicateur de densité relative

LiDAR_Penetration_Ratio - Coefficient de pénétration

Interprétation : Capacité de pénétration du LiDAR à travers la canopée

B.3.4 Indices spécialisés

CRR - Canopy Relief Ratio

Interprétation : Position relative du centre de masse vertical (0-1)

Understory_Structural_Complexity - Complexité sous-bois

Strate : 0.5-5m | *Interprétation* : Coefficient de variation du sous-bois

Forest_Edge_Density_per_m2 - Densité de lisière

Résolution : Raster 1m × 1m *interprétation** : Cet indicateur calcule la proportion de pixels présentant une variation de hauteur supérieure à 10 m dans leur voisinage immédiat. Cette métrique, initialement conçue comme proxy de densité de lisières, capture principalement la rugosité générale du couvert forestier plutôt que les vraies transitions écologiques. Sa sélection fréquente reflète probablement son rôle d'indicateur de complexité structurelle globale.

B.4 Métriques de complexité structurelle (n = 4)

B.4.1 Rugosité 3D

Rumple_Index - Indice de rugosité

Résolution : Raster 1m × 1m *Interprétation* : Ratio surface 3D/surface 2D projetée (≥ 1)

B.4.2 Diversité verticale

Foliage_Height_Diversity_Shannon - FHD Shannon

Résolution : 10 strates verticales équidistantes *Interprétation* : Diversité de Shannon de la distribution verticale

Vertical_Complexity_Index - VCI

Résolution : 10 strates verticales *Interprétation* : Proportion de strates occupées (0-1)

B.5 Métriques de couverture (n = 4)

B.5.1 Ouverture de canopée

Gap_Fraction_Ratio - Fraction de gaps

Résolution : Raster 1m × 1m *Interprétation* : Proportion de pixels < 2m de hauteur

Canopy_Openness_pct - Ouverture canopée (%)

Résolution : Raster 1m × 1m *Interprétation* : Gap fraction en pourcentage (plafonné à 100%)

B.5.2 Couvertures par strates

Canopy_Cover_Understory_pct - Couverture sous-bois

Strate : 0.5-5m | *Résolution* : Raster densité 1m *Interprétation* : Pourcentage de pixels avec végétation dans la strate

Canopy_Cover_Midstory_pct - Couverture étage moyen

Strate : 5-15m

Canopy_Cover_Canopy_pct - Couverture canopée

Strate : 15-50m

B.6 Optimisations et contrôles qualité

B.6.1 Gestion des valeurs aberrantes

- **Hauteurs** : Filtrées entre 0.5-60m pour végétation
- **CHM** : Valeurs < 0 ou > 60m mises à NA
- **Couvertures** : Plafonnées à 100%

B.6.2 Gestion des données manquantes

Attribution de NA en cas d'erreur

B.6.3 Validation des résultats

- **Cohérence spatiale** : Vérification des bornes géographiques
- **Cohérence écologique** : Validation des gammes de valeurs
- **Robustesse statistique** : Gestion des cas limites (parcelles sans végétation)

Annexe C : Spécifications techniques des variables de composition forestière

C.1 Paramètres généraux de traitement

Configuration du système

- **Source cartographique** : Cartographie des essences de Bolyn et al. (2022)
- **Résolution spatiale** : 10 m × 10 m (pixels d'entrée)
- **Échelle d'analyse** : Parcelles circulaires de 300m de rayon (28,3 ha)
- **Système de coordonnées** : ETRS89 / Lambert Belge 2008
- **Précision de classification** : 85,3% (coefficient Kappa : 0.82)

C.2 Nomenclature des essences forestières

Classification écologique (10 classes)

Classe 0 : **Non.ligneux** - Végétation non-ligneuse

Classe 1 : **Autres** - Autres essences ligneuses

Classe 2 : **Bouleaux** - Bouleaux (Betula spp.)

Classe 3 : **Chenes** - Chênes (Quercus spp.)

Classe 4 : **Douglas** - Douglas (Pseudotsuga menziesii)

Classe 5 : **Epiceas** - Épicéas (Picea abies)

Classe 6 : **Hetres** - Hêtres (Fagus sylvatica)

Classe 7 : **Melezes** - Mélèzes (Larix spp.)

Classe 8 : **Peupliers** - Peupliers (Populus spp.)

Classe 9 : **Pins** - Pins (Pinus spp.)

C.3 Proportions d'essences (n = 10)

Variables calculées (résolution : parcelle 300m)

Proportions par essence - Calcul direct

Interprétation : Pourcentage de couverture de chaque essence dans la parcelle

C.4 Indices de diversité spécifique (n = 6)

Richesse spécifique

Richesse_Essences - Nombre d'essences présentes

Interprétation : Nombre total d'essences détectées (0-10)

Signification écologique : Complexité taxonomique de l'habitat

Diversité de Shannon

Indice_Diversite_Shannon - Diversité H'

Formule : $H' = -\sum(p_i \times \ln(p_i))$ *Interprétation* : Mesure de l'incertitude taxonomique ($0-\ln(10)$) *Propriétés* : Maximum quand toutes les essences sont équiprobables

Indice_Uniformite_Shannon - Équitabilité J'

Formule : $J' = H'/\ln(S)$ *Interprétation* : Degré d'équirépartition des essences (0-1)

Diversité de Simpson

Indice_Diversite_Simpson - Indice D

Probabilité que deux individus soient de la même espèce

Indice_Diversite_Simpson_Modifie - Indice 1-D

Probabilité que deux individus soient d'espèces différentes

Formule : $1-D = 1 - \sum(p_i^2)$ *Interprétation* : Probabilité de rencontrer des essences différentes

Indice_Uniformite_Simpson - Équitabilité de Simpson

Normalisation par le nombre d'espèces

Formule : $E = (1/D)/S$ *Interprétation* : Équirépartition selon Simpson (0-1)

C.5 Variables de structure forestière (n = 2)

Dominance forestière

Pourcentage_Essence_Dominante - Dominance

Identification de l'essence la plus abondante (excluant Non ligneux)

Interprétation : Pourcentage de l'essence ligneuse dominante *Signification écologique* : Degré de mono-dominance vs. mélange

Hétérogénéité compositionnelle

Coefficient_Variation_Heterogeneite - CV des proportions

Coefficient de variation des proportions d'essences

Formule : $CV = \sigma/\mu$ Interprétation : Variabilité relative de la composition Signification écologique : Hétérogénéité du mélange forestier

Annexe D : Spécifications techniques des variables paysagères

D.1 Paramètres généraux de traitement

Configuration du système

- **Source cartographique** : WAL_UTS_2018 (Service public de Wallonie et al., 2020)
- **Résolution spatiale** : 1 m (données vectorielles par parcelle cadastrale)
- **Échelle d'analyse** : Parcelles circulaires de 300m de rayon (28,3 ha)
- **Système de coordonnées** : ETRS89 / Lambert Belge 2008
- **Couverture temporelle** : Année 2018
- **Standard de classification** : Définition INSPIRE
- **Variables finales utilisées** : 33 variables paysagères pour la modélisation

D.2 Classification écologique des usages du sol

Regroupement fonctionnel (15 catégories principales)

La nomenclature détaillée WALOUS (>50 classes) a été regroupée selon des critères écologiques pertinents pour l'avifaune :

FORÊTS

Forêt de sylviculture

"Sylviculture" -> "Foret_Sylviculture"

"Forêt de feuillus" -> "Foret_Sylviculture"

"Forêt de résineux" -> "Foret_Sylviculture"

"Forêt mixte" -> "Foret_Sylviculture"

Sapinières spécialisées

"Sapins de Noël et plants forestiers" -> "Foret_Sapins_Noel"

AGRICULTURE

Agriculture intensive

"Culture non irriguée ou irriguée" -> "Agriculture_Intensive"

"Verger et culture permanente" -> "Agriculture_Intensive"

Agriculture extensive

"Pâturage permanent" -> "Agriculture_Extensive"

"Prairie et culture fourragère" -> "Agriculture_Extensive"

Prairies

"Prairie permanente et surface naturelle" -> "Prairie"

ZONES NATURELLES

Aquatiques

"Zone naturelle aquatique" -> "Zone_Naturelle_Aquatique"

"Cours d'eau" -> "Zone_Naturelle_Aquatique"

"Plan d'eau naturel" -> "Zone_Naturelle_Aquatique"

Terrestres

"Zone naturelle terrestre" -> "Zone_Naturelle_Terrestre"

"Lande et broussaille" -> "Zone_Naturelle_Terrestre"

Non définies

"Zone naturelle non définie" -> "Zone_Naturelle_Non_Definie"

RÉSIDENTIEL (Gradient de densité)

Très faible densité

"Usage résidentiel permanent à très faible densité d'habitat" -> "Residentiel_Tres_Faible_Densite"

Faible densité

"Usage résidentiel permanent à faible densité d'habitat" -> "Residentiel_Faible_Densite"

"Usage résidentiel avec d'autres usages compatibles à faible densité" -> "Residentiel_Faible_Densite"

Densité moyenne

"Usage résidentiel permanent à densité moyenne d'habitat" -> "Residentiel_Densite_Moyenne"

"Usage résidentiel avec d'autres usages compatibles à densité moyenne" -> "Residentiel_Densite_Moyenne"

Forte densité

"Usage résidentiel permanent à forte densité d'habitat" -> "Residentiel_Forte_Densite"

"Usage résidentiel avec d'autres usages compatibles à forte densité" -> "Residentiel_Forte_Densite"

ANTHROPIQUE

Transport et infrastructure

"Transport routier" -> "Anthropique_Transport_Industrie_Services"

"Transport ferroviaire" -> "Anthropique_Transport_Industrie_Services"

"Transport par voie navigable" -> "Anthropique_Transport_Industrie_Services"

"Transport aérien" -> "Anthropique_Transport_Industrie_Services"

Services et industrie

"Services commerciaux" -> "Anthropique_Transport_Industrie_Services"

"Services culturels" -> "Anthropique_Transport_Industrie_Services"

"Services d'enseignement" -> "Anthropique_Transport_Industrie_Services"

"Services de santé et d'action sociale" -> "Anthropique_Transport_Industrie_Services"

"Industrie extractive" -> "Anthropique_Transport_Industrie_Services"

ZONES PROBLÉMATIQUES

Usages indéterminés

"Usage inconnu non cadastré" -> "Usage_Inconnu_Non_Cadastre"

"Usage inconnu cadastré" -> "Usage_Inconnu_Cadastre"

"Zones abandonnées" -> "Zone_Abandonnee"

Récréatif

"Zone récréative" -> "Zone_Recreative"

D.3 Variables de couverture d'habitat (n = 14)

Algorithme de calcul des pourcentages

Fonction d'agrégation par parcelle et catégorie

Agrégation par catégorie écologique

Variables forestières

Couverture_Foret_Sylviculture - Forêts de production

Interprétation : Pourcentage de forêts gérées (feuillus, résineux, mixtes) *Signification écologique* : Habitats forestiers structurés, ressources alimentaires prévisibles

Couverture_Foret_Sapins - Sapinières de Noël

Interprétation : Pourcentage de plantations de sapins de Noël *Signification écologique* : Monocultures temporaires

Couverture_Forestiere_Totale - Total forestier

Interprétation : Pourcentage total d'habitats forestiers *Signification écologique* : Disponibilité globale d'habitats arborescents

Variables agricoles

Couverture_Agriculture_Intensive - Agriculture intensive

Interprétation : Cultures, vergers, production intensive *Signification écologique* : Habitats homogènes

Couverture_Agriculture_Extensive - Agriculture extensive

Interprétation : Pâturages, cultures fourragères *Signification écologique* : Mosaïques agricoles, habitats semi-naturels

Couverture_Prairie - Prairies permanentes

Interprétation : Prairies permanentes et surfaces enherbées *Signification écologique* : Habitats ouverts stables, insectes abondants

Variables de naturalité

Couverture_Zones_Naturelles_Totale - Zones naturelles

Interprétation : Milieux naturels aquatiques et terrestres *Signification écologique* : Habitats refuges, biodiversité élevée

Couverture_Naturelle_Totale - Total naturel

Interprétation : Pourcentage total d'habitats naturels *Signification écologique* : Capacité d'accueil des espèces spécialisées

Variables résidentielles (gradient de densité)

Couverture_Residentiel_Tres_Faible - Très faible densité

Interprétation : Habitat rural dispersé, jardins étendus *Signification écologique* : Mosaïque semi-naturelle, espèces adaptables

Couverture_Residentiel_Faible - Faible densité

Interprétation : Zones pavillonnaires, jardins privés *Signification écologique* : Habitats anthropiques structurés

Couverture_Residentiel_Moyenne - Densité moyenne

Interprétation : Zones urbaines denses, espaces verts réduits *Signification écologique* : Contraintes spatiales, espèces urbaines

Couverture_Residentiel_Forte - Forte densité

Interprétation : Centres urbains, très forte artificialisation

Couverture_Residentiel_Totale - Total résidentiel

D.4 Variables d'anthropisation (n = 4)

Indices d'artificialisation

Couverture_Anthropique_Total - Total anthropique

Interprétation : Transport, industrie, services, infrastructures *Signification écologique* : Habitats fortement perturbés, effets de lisière

Couverture_Usage_Inconnu_Total - Usages indéterminés

Interprétation : Zones non cadastrées ou mal définies *Signification écologique* : Potentiels habitats refuges ou perturbés

Couverture_Zone_Abandonnee - Zones abandonnées

Interprétation : Friches, zones délaissées *Signification écologique* : Succession écologique, recolonisation

Indice_Artificialisation - Gradient d'artificialisation

Formule : $(\text{Artificiel}) / (\text{Naturel}) \times 100$ *Interprétation* : Ratio artificialisation/naturalité *Signification écologique* : Gradient de pression anthropique

D.5 Indices de configuration paysagère (n = 9)

Superficie_Totale_Ha - Superficie analysée

Interprétation : Superficie totale couverte par l'analyse *Contrôle qualité* : Doit approcher 28,3 ha (théorique)

Fragmentation_Totale - Nombre total de fragments

Interprétation : Nombre total de polygones dans la parcelle *Signification écologique* : Complexité spatiale, effets de lisière

Richesse_Habitat - Diversité des types d'habitat

Interprétation : Nombre de types d'habitats différents *Signification écologique* : Hétérogénéité des niches disponibles

- **Indice_Division_Paysage** : division spatiale du paysage
- **Indice_Maille_Effective** : connectivité effective
- **Indice_Fragmentation_Split** : degré de fragmentation
- **Indice_Cohesion_Paysage** : cohésion spatiale
- **Indice_Agregation_Paysage** : agrégation des éléments
- **Indice_Forme_Paysage** : complexité des formes

D.6 Indices synthétiques d'habitat (n = 6)

Variables de dominance et diversité

Type_Dominant - Habitat dominant

Interprétation : Type d'habitat le plus représenté *Signification écologique* : Caractérisation primaire du paysage

Pourcentage_Dominant - Dominance relative

Interprétation : Pourcentage du type d'habitat dominant *Signification écologique* : Degré d'homogénéité vs hétérogénéité

Shannon_Habitat - Diversité paysagère

Formule : $H' = -\sum(\pi_i \times \ln(\pi_i))$ *Interprétation* : Diversité de Shannon des habitats *Signification écologique* : Complexité de la mosaïque paysagère

Indices spécialisés pour l'avifaune

Indice_Naturalite - Gradient de naturalité

Interprétation : Proportion d'habitats naturels *Signification écologique* : Qualité globale pour espèces spécialisées

Indice_Habitat_Forestier - Spécialisation forestière

Interprétation : Proportion d'habitats forestiers *Signification écologique* : Disponibilité pour espèces forestières

Indice_Mosaïque_Agricole - Complexité agricole

Interprétation : Proportion d'agriculture extensive et prairies *Signification écologique* : Habitats semi-naturels, espèces de milieux ouverts

Indice_Pression_Urbaine - Pression anthropique

Interprétation : Intensité de l'urbanisation *Signification écologique* : Pression sur les espèces sensibles

Annexe E : 4.11.4 Limitations des tests statistiques

L'approche statistique adoptée souffre d'une inadéquation fondamentale entre les objectifs écologiques de l'étude et le cadre inférentiel classique. Le test de DeLong répond à la question "la différence d'AUC est-elle statistiquement significative ?" plutôt qu'à la question écologiquement pertinente "cette différence représente-t-elle un gain prédictif substantiel pour la gestion conservatoire ?". Cette confusion entre significativité statistique et importance pratique constitue un biais majeur.

Violations des hypothèses du test de DeLong :

Le test de DeLong assume que les courbes ROC suivent une distribution asymptotiquement normale, hypothèse rarement vérifiée pour les espèces à faible prévalence (< 10%) représentant 60% de l'échantillon. L'algorithme exige également que les modèles comparés soient testés sur les mêmes observations exactes, condition problématique dans un contexte de validation croisée spatiale où les prédictions résultent d'agrégations sur

différents folds. La covariance (Cov_{12}) entre modèles partageant des variables communes introduit une dépendance non prise en compte dans l'interprétation des p-values.

Sensibilité excessive à la taille d'échantillon :

Le test présente une sensibilité paradoxale à la taille d'échantillon : pour les espèces communes ($n > 500$), des différences écologiquement triviales ($\Delta AUC < 0.01$) peuvent atteindre la significativité statistique, tandis que pour les espèces rares ($n < 100$), des différences substantielles ($\Delta AUC > 0.05$) peuvent rester non-significatives. Cette dépendance à la taille d'effet standardisée rend les comparaisons inter-espèces problématiques.

Conservatisme excessif de la correction de Bonferroni :

La correction de Bonferroni assume l'indépendance complète des tests multiples, hypothèse violée lorsque les modèles partagent des sous-ensembles de variables. Cette correction augmente artificiellement le risque d'erreur de type II (faux négatifs), particulièrement problématique pour les espèces rares où la puissance statistique est déjà limitée. Des approches alternatives comme la False Discovery Rate (FDR) de Benjamini & Hochberg (1995) seraient plus appropriées pour contrôler le taux de fausses découvertes tout en préservant la puissance.

Inadéquation du bootstrap standard :

L'implémentation d'un bootstrap simple ($n = 1000$) ignore la structure spatiale des données, violant l'hypothèse d'échangeabilité des observations. Un bootstrap spatial par blocs serait nécessaire pour respecter l'autocorrélation spatiale détectée dans les résidus. Le nombre d'itérations (1000) peut s'avérer insuffisant pour une estimation robuste des intervalles de confiance, particulièrement pour les métriques de queue de distribution comme les quantiles extrêmes.

Seuils de classification arbitraires et non-validés :

Les seuils définissant les catégories de gain LiDAR (0.01, 0.02, 0.05) reposent sur des conventions statistiques plutôt que sur des critères écologiques validés. Aucune méta-analyse ou étude de calibration n'établit qu'un $\Delta AUC = 0.05$ correspond effectivement à un "gain substantiel" en termes de capacité prédictive opérationnelle. Cette classification binaire ignore la continuité des effets et peut masquer des patterns écologiques plus subtils.

Alternative méthodologique recommandée :

L'analyse statistique a été réalisée dans un laps de temps assez restreint et mériterait un approfondissement, cette partie constituant un domaine extrêmement vaste et complexe qui demande une véritable expertise.