

## Identification des espèces d'oiseau indicatrices des trames du réseau écologique nichant en wallonie et délimitation des zones et des corridors de ce réseau

**Auteur :** Pollet, Thomas

**Promoteur(s) :** Dufrêne, Marc

**Faculté :** Gembloux Agro-Bio Tech (GxABT)

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

**Année académique :** 2019-2020

**URI/URL :** <http://hdl.handle.net/2268.2/10205>

---

### 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.

---

**IDENTIFICATION DES ESPECES D'OISEAUX NICHANT EN WALLONIE INDICATRICES DES  
TRAMES DU RESEAU ECOLOGIQUE ET DELIMITATION DES ZONES CŒURS DE  
BIODIVERSITE ET DES CORRIDORS DE CE RESEAU**

**POLLET THOMAS**

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

**ANNÉE ACADÉMIQUE 2019-2020**

**PROMOTEUR: PR MARC DUFRÊNE**





*Copyright © 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<sup>1</sup> de Gembloux Agro-Bio Tech.*

*" Le présent document n'engage que son auteur "*

---

<sup>1</sup> Dans ce cas, l'autorité académique est représentée par le(s) promoteur(s) membre du personnel(s) enseignant de GxABT



**IDENTIFICATION DES ESPECES D'OISEAUX NICHANT EN WALLONIE INDICATRICES DES  
TRAMES DU RESEAU ECOLOGIQUE ET DELIMITATION DES ZONES CŒURS DE  
BIODIVERSITE ET DES CORRIDORS DE CE RESEAU**

**POLLET THOMAS**

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

**ANNÉE ACADÉMIQUE 2019-2020**

**PROMOTEUR: PR MARC DUFRÊNE**



## Remerciements

En tant qu'auteur de ce travail de fin d'étude, je souhaiterais remercier tout d'abord l'ASBL Natagora et plus précisément Mr Jean-Yves Paquet et Mr Antoine Derouaux ainsi que le DEMNA pour le partage de leurs données d'observations et le traitement de celles-ci. Ils m'ont également donné des conseils très précieux pour la réalisation de ce travail.

Je remercie cordialement Mr Axel Bourdouxhe et Mr. Thomas Coppée pour l'incalculable aide qu'ils ont fournie à la réalisation de ce travail ainsi que l'entière de l'unité Biodiversité et Paysage pour leur accueil chaleureux et leur convivialité à toutes épreuves.

Pour finir, il est aussi très important de remercier le promoteur de ce travail, Pr Marc Dufrêne pour les révisions, conseils et aides générales qui ont été fournis ainsi que les membres du jury de la défense orale, Pr Arnaud Monty, Pr Grégory Mahy, Pr Yves Brostaux et une fois de plus, Mr Jean-Yves Paquet.



## Abstract

Due to a worrying loss of many bird species in the Walloon Region over the last few decades, it is imperative to develop strong and efficient protection measures to put an end to this phenomenon. One of the most effective tools in biodiversity conservation is with no doubt the construction of an ecological network. The aim of this work is to initiate the first steps in getting an operational ecological network on the scale of the Walloon region and, in this case, focusing on birds. These initial steps consist of the realisation of a first proposal of an ecological network composed of biodiversity core areas which are classified by bio geographical regions and by ecological frame. This term defines a set of similar biotopes which hosts species groups with similar natural needs. Essentially, this is an exploratory work about the construction of an ecological network on the scale of the entire Walloon Region.

In order to do this, abundance models of about sixty species breeding in Belgium have been provided. They have been constructed based on the GAM algorithm and the presence/absence data from the European Atlas for birds' inventories. For the rest of the species breeding in Wallonia, their filtered observation data has been provided by Natagora and has come from sensor platforms of the OFFH and "Observation.be". The presence data is then used to model the potential ecological niche of each species. The MaxEnt algorithm has been tested three times, once with the default background (pseudo-absences) then twice with different improved methods compared to the first one. Thanks to the latter ones, the GAM algorithm has also been able to be tested. The environmental data has been provided by an ecotope layer specially measured to fit the kilometric squares of the European atlas. The final selected method, which offers the best results, is the one that uses MaxEnt with a random selection of 10 000 pseudo-absence points. The two model types obtained at the end are all on the one square kilometre scale.

In addition, the modelled presences are used with the land cover variables from the ecotope layer which each describe a frame to make a redundancy analysis which in turn allows the determination of the indicator species. Subsequently, a potential ecological network is built based on the indicator species presences, the alpha index measured on the modelled presences and on the same index but weighted by the RDA results. The obtained results indicate a succession of kilometric squares chosen as a core area for each distinct frame and on the entire Walloon Region and they cover 25% of it. Finally, several analysis' have been conducted to characterize and assess these networks and then completed by the usage of indexes. Broadly speaking, the ecological network is quite convincing seeing the value of the different indexes. Indeed, the network majority represents their own frame well and contain kilometric squares with the best diversity indexes. However, some modifications and assessments could be brought to this exploratory work methodology.

**Keywords:** Ecological network, birds, indicator species, modelling, MaxEnt, Walloon Region, ecological frame, citizen science.

## Résumé

En raison de la perte inquiétante de nombreuses espèces d'oiseaux en Région Wallonne ces dernières décennies, il est impératif de développer des mesures de protection fortes et efficaces afin d'endiguer ce phénomène. L'un des outils les plus performants en matière de conservation de la nature est sans doute la construction d'un réseau écologique. L'objectif de ce travail est donc d'initier les premiers pas vers l'obtention d'un réseau écologique opérationnel à l'échelle de la Wallonie centré dans ce cas, sur les oiseaux. Ces premiers pas consistent en la réalisation d'une première proposition d'un réseau écologique fonctionnel composé de zones cœurs de biodiversité déterminées par région biogéographique et par trame écologique. Ce terme définit un ensemble de biotopes semblables abritant des groupes d'espèces ayant des besoins naturels similaires. Il s'agit donc d'un travail exploratoire sur la construction d'un réseau écologique à l'échelle de l'entière de la Région wallonne.

Pour se faire, les modèles d'abondance d'une soixantaine d'espèces d'oiseaux nichant en Belgique ont été fournis. Ils ont été construits sur base de l'algorithme de GAM et des données de présence/absence issues des inventaires de l'Atlas européen des oiseaux. Pour le reste des espèces nichant en Wallonie, leurs données d'observation filtrées ont été fournies par Natagora et proviennent des plateformes de recensement de l'OFFH et « Observation.be ». Ces données de présences sont utilisées afin de modéliser la niche écologique potentielle de chacune de ces espèces. L'algorithme MaxEnt est testé à trois reprises avec un background (pseudo-absence) par défaut et avec deux méthodes d'améliorations de celui-ci. Via ces dernières, l'algorithme de GAM a pu également être testé. Les données environnementales sont fournies par une couche écotope spécialement mesurée pour convenir aux carrés kilométriques de l'Atlas européen. La méthode finalement choisie et qui offre les meilleurs résultats est celle utilisant MaxEnt avec une sélection aléatoire de 10 000 points de pseudo-absences. Les deux types de modèles finalement obtenus sont tous à l'échelle de 1km<sup>2</sup>.

Ensuite, les présences modélisées obtenues sont utilisées avec les variables d'occupation du sol de la couche écotope qui décrivent chacune une trame pour réaliser des analyses de la redondance permettant la détermination des espèces indicatrices de ces trames. Après cela, un réseau écologique potentiel est construit en se basant sur la présence des espèces indicatrices, l'indice alpha mesuré, les présences modélisées et sur ce même indice mais pondéré avec les résultats des RDA. Les résultats obtenus offrent une succession de carrés kilométriques désignés en tant que zone cœur de biodiversité par trame écologique distincte sur l'entière de la Région wallonne recouvrant 25% de celle-ci. Pour finir, une série d'analyse a été réalisée pour les caractériser et les évaluer, complétée par l'utilisation de divers indices. D'une manière générale, les réseaux écologiques ainsi délimités sont assez convaincants en vue de la valeur de ces indices. En effet, la majorité des réseaux représente bien leur trame et contient les carrés kilométriques ayant les meilleurs indices de diversité. Cependant, quelques modifications et améliorations restent tout de même à être apportées à cette méthodologie de travail exploratoire.

**Mots clés :** Réseau écologique, oiseaux, espèces indicatrices, modélisation, MaxEnt, Région Wallonne, trame écologique, plateforme de recensement.



# Table des matières

Remerciements .....	1
Abstract .....	2
Résumé.....	3
Table des matières .....	5
Liste des tableaux .....	7
Table des figures .....	8
I. Introduction .....	9
1.1. Acquisition et prétraitement des données .....	14
1.2. Modélisation .....	14
1.2.1. Maxent.....	15
1.2.2. GLM et GAM.....	16
1.2.3. Sélection des variables .....	17
1.2.4. Biais d'échantillonnage.....	18
1.2.5. Sélection du Background (pseudo-absence) .....	19
1.2.6. Évaluation des modèles .....	21
1.2.7. Seuil de probabilité.....	22
1.2.8. Autres critères d'évaluation .....	22
1.3. Identification des espèces indicatrices.....	23
1.4. Réseau écologique .....	23
1.5. Évaluation du réseau écologique .....	24
II. Matériel et méthode.....	27
2.1. Acquisition des données d'observation .....	27
2.2. Gestion des données .....	27
2.3. Analyse des données de présence brutes .....	28
2.4. Modélisation .....	29
2.4.1. Sélection des variables .....	29
2.4.2. Création du background (Pseudo-absences).....	29
2.4.3. Sélection de la méthode.....	30
2.4.4. Modélisation par MaxEnt.....	30
2.4.5. Évaluation et validation des modèles .....	31
2.4.6. Comparaison et choix de la méthode finale .....	32
2.5. Sélection des espèces indicatrices .....	32
2.6. Réseau écologique existant.....	33
2.7. Réseau écologique potentiel .....	34
2.8. Analyse des réseaux écologiques .....	35

III.	Résultats .....	39
3.1.	Révisions des trames écologiques .....	39
3.2.	Modélisation .....	40
3.3.	Sélection des espèces indicatrices .....	44
3.4.	Réseau écologique existant.....	47
3.5.	Réseau écologique potentiel .....	50
3.6.	Analyses supplémentaires.....	55
IV.	Discussion .....	61
4.1.	Acquisition et traitement des données .....	61
4.2.	Révision des trames .....	61
4.3.	Modélisation .....	62
4.3.1.	Méthode proposée .....	62
4.3.2.	Indice et méthode de sélection .....	63
4.3.3.	Modélisation finale.....	66
4.4.	Construction des réseaux écologiques .....	67
4.4.1.	Sélection des espèces indicatrices .....	67
4.4.2.	Délimitation des zones cœurs.....	68
4.4.3.	Délimitations des corridors .....	73
4.5.	Analyse des réseaux.....	73
4.6.	Analyses complémentaires .....	75
V.	Conclusion.....	77
5.1.	Méthodes de modélisation .....	77
5.2.	Modélisation .....	77
5.3.	Construction des réseaux écologiques .....	78
VI.	Bibliographie.....	79

## Liste des tableaux

Tableau 1: Liste des trames thématiques présentées par Dufrêne et al., 2019 ainsi que leur définition sommaire et les habitats WalEUNIS correspondants. ....	11
Tableau 2: Listes des trames associées aux variables d'occupation du sol (écotope) les caractérisant via leur code ainsi que leur signification. ....	28
Tableau 3: Liste des méthodes testées pour la sélection des pseudo-absences (background) lors de l'utilisation de MaxEnt. ....	30
Tableau 4: Indices d'évaluation des modèles sur base d'une matrice de confusion ainsi que leurs formules. ....	31
Tableau 5: Indice d'évaluation d'un réseau associé à leur formule et leurs bornes supérieures et inférieures. ....	35
Tableau 6: Liste des trames candidates à la composition du réseau écologique sur l'avifaune associées à leur fréquence relative ....	40
Tableau 7: Liste des variables écotopes utilisées pour la modélisation classée selon leur différent type. ....	40
Tableau 8: Résultats des trois méthodes de sélection du background concernant la modélisation via l'algorithme MaxEnt. ....	41
Tableau 9: Valeur des AUC des modèles construits via l'algorithme de GAM. ....	42
Tableau 10: Liste des espèces dont la qualité de leur modèle est jugée insuffisante. ....	43
Tableau 11: Tableau reprenant les $\cos^2$ de chaque espèce ayant au moins une valeur supérieure au seuil (10%) dans la région des Ardennes et concernant les espèces rares.. ....	46
Tableau 12: Tableau récapitulatif reprenant les espèces indicatrices finalement sélectionnées pour chaque trame au sein de chaque région précédemment délimitée. ....	46
Tableau 13: Nombre d'espèces seuil différenciant des autres, un carré kilométrique ayant une biodiversité observée remarquable. ....	48
Tableau 14: Seuils appliqués aux cartes rasters afin de déterminer sous quel statut de protection un carré kilométrique peut être assimilé. ....	49
Tableau 15: Seuils de différenciation imposés à la valeur de l'indice alpha pondéré par carré kilométrique par région et pour chaque trame écologique. ....	50
Tableau 16: Seuils du nombre d'espèces imposés aux carrés kilométriques pour leur sélection en zone cœur des cinq régions. ....	50
Tableau 17: Résultats des indices de probabilité de connexion et d'Harary pour les réseaux de chaque trame ainsi que pour les trois distances de dispersion. ....	53
Tableau 18: Valeurs moyennes des indices par distance de dispersion ainsi que leur écart-type. ....	53
Tableau 19: Tableau reprenant la proportion de surface partagée entre deux réseaux. Le sens de lecture du tableau à deux entrées est indiqué par la flèche. ....	55
Tableau 20: Tableaux des mesures de la précisions des réseaux écologiques. ....	57
Tableau 21: Tableaux récapitulatifs des moyennes ainsi que de leurs écarts-types des indices alpha observé et modélisé. ....	58
Tableau 22: Tableau de contingence montrant le nombre de zones cœurs populaires pour les ornithologues et permettant le test $\chi^2$ . ....	59
Tableau 23: Valeurs attendues si les deux variables testées étaient vraiment indépendantes. ....	59
Tableau 24: Valeur des corrélations entre l'indice alpha observé, alpha modélisé et le nombre d'observations par carré kilométrique avec l'occupation du sol. ....	59
Tableau 25: Surface des sites mis sous statut de protection et couverts par le réseau global. ....	60
Tableau 26: Statistiques descriptives des valeurs des indices alpha suite à la modélisation et sur base des données empiriques. ....	60

## Table des figures

Figure 1: Schéma reprenant les points essentiels de la méthodologie de sélection des espèces indicatrices, de désignation des zones cœurs de biodiversité ainsi que de la délimitation des réseaux écologiques.....	37
Figure 2: Schéma reprenant les analyses principales réalisées après la construction des différents réseaux .....	38
Figure 3: Trames écologiques redéfinies afin de mieux englober les demandes écologiques, abiotiques et biotiques de l'avifaune en Région Wallonne. ....	39
Figure 4: Cartes des présences/absences résultantes de la modélisation de deux espèces via les trois méthodes différentes. ....	42
Figure 5: Couche raster représentant les probabilités "MaxEnt" de présence et les présences et absences modélisées de trois espèces différentes choisies selon la qualité de leur modèle respectif. ....	44
Figure 6: Cartographie des territoires biogéographiques considérés séparément lors de l'analyse des espèces indicatrices. ....	44
Figure 7: Analyse de la redondance des données de présences prises en Ardenne et au nord du sillon Sambre et Meuse/Condroz avec prise en compte des occupations du sol représentant l'entièreté des trames. ....	45
Figure 8: Représentation graphique des résultats de la RDA concernant les données en Ardenne pour la trame "Forêts feuillues" et le cercle de corrélation associé .....	45
Figure 9: Graphiques résultants des RDA en Lorraine et Nord sillon/Condroz pour la variable de milieux ouverts humides ainsi que le cercle de corrélation associé à chaque graphe .....	47
Figure 10: Cartes des proportions surfaciques au sein d'un km <sup>2</sup> des différents statuts relatif à la conservation de la nature en Belgique .....	48
Figure 11: Représentation cartographique de la somme des surfaces par carré kilométrique pondérées des sites sous statut et le résultat de l'application des seuils sur l'indice alpha observé.....	49
Figure 12: Ensemble des zones cœurs existantes selon la diversité observée et les statuts de protection ou autres dénominations relatives à la conservation de la biodiversité. ....	49
Figure 13: Couche raster représentant l'indice alpha mesuré à l'aide des présences modélisées de toutes les espèces et celui pondéré par le résultat des RDA pour la trame forestière.....	50
Figure 14: Ensemble des liens pouvant exister entre les différentes taches d'habitat favorable selon 3 distances de dispersion moyennes et ceci, pour la trame des milieux forestiers feuillues et des cultures. ....	51
Figure 15: Représentation cartographique du réseau écologique final de la trame des forêts résineuses et des prairies.....	51
Figure 16: Comparaison cartographique des corridors représentés via une carte de chaleur avec et sans pondération selon le dPC connector des liens .....	52
Figure 17: Carte représentant les carrés kilométriques repris en tant que zone cœur ainsi que la fréquence correspondante et le graphique représentant la fréquence des valeurs de pixels de cette couche raster.....	56
Figure 18: Cartographie des zones cœurs de la trame des prairies en Ardenne via la première méthode et la méthode alternative. ....	70

## I. Introduction

La relation entre l'homme et le paysage peut se résumer le plus fréquemment à la fragmentation excessive des habitats naturels, à leur isolation et leur érosion (Melin, 1997a; Bernier et al., 2013; Lightbown et al., 2013). Ceci mène directement à une diminution globale et non négligeable de la biodiversité via la dégradation des processus naturels à l'échelle du paysage ce qui décroît également les services écosystémiques rendus par ce dernier (Jongman, 2002; Mander et al., 2003; Jongman et al., 2011). En effet, la fragmentation cause une diminution de la disponibilité en habitat, défavorise la bonne dispersion des espèces (Avon et al., 2016) et est considérée comme la cause principale de l'actuelle érosion de la biodiversité (Fahrig, 2003). Ce phénomène est d'autant plus vrai en Europe occidentale et donc en Belgique due à une densité démographique plus importante et à de nombreuses pratiques non respectueuses de l'environnement. C'est au sein de cette dernière que l'un des taux de fragmentation les plus élevés a pu être mesuré (Jaeger et al., 2016). Or, l'un des plus grands challenges que l'humanité doit surmonter à présent est de subvenir aux besoins des populations actuelles et futures via le développement durable. Celui-ci peut s'obtenir uniquement si le paysage parvient à supporter les demandes écologiques, sociales et économiques de l'Homme (Opdam et al., 2006). Le défi est donc de maintenir les ressources clés du paysage dans une dynamique évolutive rapide que connaît la société actuelle (Jaeger et al., 2016). L'outil existant à ce jour qui semble être le plus efficace pour lutter contre ce phénomène et permettant de relever ce défi serait l'élaboration d'un réseau écologique au sein de ces milieux fragmentés (Melin, 1997). En effet, la simple protection de grandes zones à haute biodiversité n'a que très rarement fonctionné correctement. Un flux constant entre différentes populations doit être maintenu afin de minimiser les risques liés à l'isolation. Ceux-ci sont principalement la dégénérescence génétique provoquée par la consanguinité, les variations démographiques aléatoires, la perte de diversité génétique et l'affaiblissement face aux maladies (Hansson et al., 1992). Dans la même logique, le manque de connexions empêche les espèces de parvenir à migrer lors de fortes perturbations au sein de leur habitat (Sawyer et al., 2011) et d'éventuellement recoloniser une zone récemment abandonnée (Bourdouxhe et al., 2019).

Un réseau écologique est défini par la Région Wallonne comme un ensemble de biotopes pouvant offrir aux espèces un habitat répondant à leurs demandes écologiques dans une optique de maintien à long terme (Delescaillle, 1993). Il est donc constitué de zones centrales de conservation ou zones cœurs de biodiversité. À celles-ci s'ajoutent les zones tampons ou zones de développement se retrouvant autour d'elles et qui possèdent un statut de protection légèrement moins fort. Ces ensembles sont reliés par des corridors ou couloirs naturels caractérisés par une forte concentration en éléments paysagers du maillage écologique (Dufrêne, 2005; Jongman et al., 2011). Ces derniers permettent aux espèces concernées par le réseau de s'y déplacer librement et facilement. Grâce à cela, un réseau maintient la biodiversité des milieux naturels, la diversité au sein même des populations et la diversité des interactions au sein des communautés (Devillers et al., 2003). Cependant, il y a une différence nette entre le réseau et le maillage écologique. Ce dernier correspond aux éléments paysagers faisant partie du réseau écologique qui sont par exemple des haies, des buissons ou des alignements d'arbres. Il s'agit donc des éléments du paysage permettant le déplacement des espèces dans un corridor.



Cette thématique a commencé à se développer dans les années 80 suite à la théorie de la biogéographie des îles (MacArthur et al., 1967) et à l'émergence de l'écologie du paysage (Forman et al., 1986) avec le concept des métapopulations (Levins, 1969). Elle a depuis, pris beaucoup d'importance. Les maîtres en la matière semblent être les Allemands qui ont débuté leur programme il y a 25 ans (Jaeger et al., 2016) et les Suisses qui ont également bien développé leur réseau national. La France a suivi en élaborant son système de trame verte et bleue (Amsallem et al., 2013; Locquet et al., 2018). Les Pays-Bas ont également un réseau assez robuste qui fut initié en 1989 et inclus dans leur programme de développement rural (Jongman et al., 2004, 2011). Pour finir, la Flandre a incorporé le concept de réseau écologique dans sa politique de conservation via la Structure verte principale (De Blust et al., 1992). Ainsi, de nombreux pays et régions tendent à créer des liaisons écologiquement pertinentes entre les taches naturelles rémanentes qu'ils possèdent. De cette manière, ils favorisent le maintien de la biodiversité et d'un flux génétique constant ainsi que la recolonisation d'habitats à restaurer (Jaeger et al., 2016).

Pour citer un autre réseau existant et célèbre, il y a bien entendu le réseau Natura 2000 qui s'étend sur toute l'Union Européenne et qui permet de mettre en œuvre les directives européennes oiseau et habitat (Evans, 2012). Toujours à l'échelle européenne, le projet du nom de réseau écologique PANeuropéen s'est également développé. Celui-ci est composé de trois sous projets travaillant chacun sur une zone de l'Europe distincte, à savoir, l'Europe occidentale (Jongman et al., 2006), l'Europe du Sud-est et l'Europe centrale et orientale (Jongman et al., 2011). Ici, les zones cœurs sont déterminées via la sélection d'espèces importantes selon divers critères et de l'analyse de l'occupation du sol. Les corridors, eux, sont déterminés principalement via les voies de migration, les chaînes de montagnes, les cours d'eau ou encore des successions de taches forestières. Ce programme agit directement sous l'égide du conseil européen et du programme environnemental des Nations Unies (Council of Europe, 1996). Ensuite, afin d'ajouter les pays non membres au réseau Natura 2000, le réseau Emerald et transeuropéen ont été créés dont le premier suit exactement les mêmes principes que Natura 2000 (Council of Europe, 2009). Concernant la Wallonie, le projet central mené par le gouvernement wallon se nomme la Structure Ecologique Principale et a pour objectif d'englober l'entièreté des zones à fort intérêt reconnu ou potentiel pour la biodiversité (Dufrêne, 2005). De plus, des travaux d'élaboration de réseaux écologiques spécifiques à une espèce, un groupe d'espèces ou selon les types d'habitats ont également vu le jour (Bourdouxhe et al., 2019).

Malheureusement, les efforts entrepris jusqu'ici semblent loin d'être suffisants. Il y a de nombreux problèmes et irrégularités au sein de bon nombre de ces projets. En effet, les réseaux mis en place ont eu un impact assez restreint et n'ont que peu freiné les effets de la fragmentation (Jaeger et al., 2016). Pour le réseau Natura 2000 par exemple, celui-ci possède de nombreuses lacunes administratives et structurelles même si des effets positifs sur des populations d'oiseaux ont déjà été prouvés (Pellissier et al., 2013). Les corridors entre les différentes unités de gestion semblent ne pas être adéquatement définis et l'évolution des écosystèmes est mal prise en compte qu'elle soit intrinsèque au milieu ou due aux changements climatiques (Evans, 2006). La combinaison de ces deux facteurs fait remonter à la surface le problème des zones protégées mais isolées car des espèces prioritaires qui s'y trouvent risquent tout de même de disparaître (Jongman et al., 2011). Ainsi ce réseau pourrait se résumer à un ensemble de zones protégées de façon inégale et non suffisamment connectées. Une infrastructure verte devrait donc idéalement se développer autour afin d'améliorer cette connexion (Green infrastructure, 2007; Jaeger et al., 2016). Concernant le réseau écologique européen, celui-ci possède le principal défaut d'avoir été créé à une échelle assez grossière.

La variabilité des sources d'informations concernant les types de sols, les bassins versants et la géologie ne leur ont également pas permis de prendre ces variables en compte réduisant intrinsèquement la pertinence du réseau. Cependant, cela offre tout de même une vision globale de la situation européenne et met déjà en avant des zones prioritaires (Jongman et al., 2011).

La situation belge et en Europe occidentale est préoccupante. L'objectif à atteindre le plus rapidement possible est la détermination des zones d'importance stratégique et des corridors. Ensuite, il faut les protéger via des moyens légaux forts qui doivent être intégralement respectés. L'un des plus gros soucis en Belgique est que non seulement un réseau écologique à proprement dit n'est pas encore opérationnel mais qu'également, les mises sous réserve de zones cœurs de biodiversité sont pratiquement absentes. Elles représentaient environ 0,5% du territoire wallon en 2005 (Dufrêne, 2005) et n'auraient augmenté que jusque 1,22% en l'espace de 12 ans (Biodiversité.wallonie.be, 2020). Tout ceci, alors que l'élaboration de ce réseau est décrite dans le plan de conservation de la biodiversité du gouvernement belge (Melin, 1997b; Dufrêne, 2004). Il y a donc des efforts à faire en matière de gestion et de maintien des espèces au sein de la Wallonie. En effet, il est très clairement stipulé dans la législation européenne mettant en œuvre la convention sur la biodiversité (Nations Unies, 1992; Commission Européenne, 2013) qu'il faut maximiser au possible les zones de protection au sein de chaque pays.

En vue de cette efficacité absolue malheureusement non acquise et du manque d'actions concrètes réalisées en Wallonie en matière de conservation à l'échelle régionale, la réalisation d'un réseau écologique reste une priorité. Le point de départ serait une analyse approfondie de toutes les données environnementales, écologiques et naturalistes acquises durant ces dernières années. Elles sont, pour la plupart, relativement complètes et bien développées, en grande partie grâce à l'augmentation du taux de relevés scientifiques faits par des citoyens (Johnston et al., 2018). Tout ceci permettrait de matérialiser un réseau fonctionnel théorique à l'échelle de toute la Wallonie en travaillant par groupe spécifique, par région biogéographique et par trames écologiques distinctes. Ces dernières sont définies comme des ensembles de biotopes semblables abritant des groupes d'espèces ayant des besoins naturels similaires (Dufrêne et al., 2019) et sont reprises avec leur définition au sein du tableau 1.

Tableau 1: Liste des trames thématiques présentées par Dufrêne et al., 2019 ainsi que leur définition sommaire et les habitats WalEUNIS correspondants (Dufrêne et al., 2005).

Trame	Définition	Biotope associé
Forestière feuillue	Recouvrement majoritaire d'arbres décidus ou en mélange avec des conifères	G1 Forêts feuillues décidues G4 Plantations mixtes feuillues-conifères
Milieux ouverts secs	Zones ouvertes à dominance herbacée, buissonnante ou légèrement arbustive et ayant un sol à capacité drainante moyenne à excessive	E1 Pelouses sèches E2 Prairies mésophiles F4 Landes des régions tempérées
Milieux ouverts humides	Zones ouvertes à dominance herbacée, buissonnante ou légèrement arbustive et ayant un sol à capacité drainante faible à très mauvaise	D1 Tourbières hautes et tourbières de couverture D2 Tourbières de vallées, bas-marais et tourbières de transition D4 Bas-marais riches alcalins D5 Roselières sèches et magnocariçaias E5 Mégaphorbiaies et ourlets forestiers
Eaux de surface	Eaux courantes ou stagnantes ainsi que la végétation aquatique associée	C1 Eaux stagnantes C2 Eaux courantes C3 Végétation aquatique
Bocages et lisières	Ensembles de haies, les zones en lisière forestière, les regroupements d'arbres isolés en prairies ainsi que les zones à dominances arbustives et buissonnantes, parcelles forestières récemment exploitées, jeunes plantations et plantations de sapins de Noël	G5 Alignements d'arbres, zones récemment exploitées, forêts et taillis aux jeunes stades F3 Fourrés tempérés et méditerranéo-montagnards F9 Fourrés riverains et fourrés sur bas-marais FA Haies FB Plantations de ligneux bas
Milieux agricoles	Zones cultivées ou récemment cultivées ou ayant un lien direct avec la production agricole comme des prairies de fauche ou des champs en jachère	I1 Cultures Ia Prairies temporaires de fauche

La Région Wallonne, quant à elle, peut être séparée en six régions distinctes pouvant influencer la biodiversité locale. Ces dernières sont le territoire hennuyer-brabançon-hesbignon, le Condroz, la Fagne-Famenne-Calestienne, l'Ardenne, la Gaume et Lorraine belge et le sillon sambro-mosan. Ils sont considérés comme des zones aux conditions écologiques moyennes homogènes ce qui signifie que la présence ou l'absence d'une espèce pourra être expliquée de la même façon sur tout le territoire. Cette classification s'est basée sur l'ensemble des biotopes et des espèces caractéristiques du territoire considéré ayant également été réalisée par Dufrêne et al., (2019). Le premier territoire est le moins boisé de tous et comporte une agriculture intensive prépondérante. Le Condroz quant à lui, est une alternance entre zones d'agriculture et bandes boisées avec peu de zones humides naturelles. Ensuite, la Fagne-Famenne-Calestienne se présente comme la dépression séparant le plateau du Condroz et le plateau Ardennais. Ce territoire est caractérisé par des affleurements rocheux d'origine calcaire ayant permis l'apparition de pelouses calcaires semi-naturelles principalement concentrées en Calestienne. Cette zone est beaucoup moins propice à l'agriculture intensive laissant place à des prairies et bocages en équilibre avec des massifs forestiers assez importants. D'un point de vue général, ce territoire est celui présentant le meilleur état de conservation de l'avifaune. L'Ardenne est considérée comme étant le territoire le plus forestier de tous, tout en possédant des milieux ouverts d'intérêt biologique important. La Lorraine belge est également très boisée avec des assises calcaires et gréseuses ayant justifié la création de carrières. Cependant, ce territoire comporte un paysage forestier et agricole fortement altéré par l'agriculture intensive et son urbanisation rapide. Finalement, le sillon sambro-mosan peut être décrit par un relief particulier avec de nombreux versants boisés ainsi que des affleurements schisteux, gréseux et calcaires et surtout avec des pelouses calaminaires impliquant une biodiversité exceptionnelle (Jacob et al., 2010).

Concernant le groupe d'espèces choisi, dans la plupart des méthodes présentées dans la littérature pour la construction d'un réseau, la sélection d'espèces indicatrices du milieu ou de la trame revient souvent (Jongman et al., 2011; Bernier et al., 2013; Locquet & Clauzel al., 2018). Ces espèces sont considérées comme représentatives d'un milieu ou d'un groupe d'espèces ayant des exigences écologiques similaires. L'intérêt de celles-ci est dans un premier temps de réduire la charge de travail et de traitement des données. Dans un second temps, les espèces sélectionnées le sont généralement pour leur caractère parapluie (Beier et al., 2008; Locquet et al., 2018) ce qui signifie que leur protection permettra simultanément celle de nombreuses autres espèces (Roberge & Angelstam, 2004). Elles peuvent également être sélectionnées selon leur niveau de vulnérabilité, leur statut légal ou encore le caractère complet de leurs données qui sont disponibles (Albert & Chaurand, 2018).

L'efficacité d'un réseau écologique se définit premièrement par ses objectifs biologiques qui sont représentés par des espèces ou des habitats cibles (Dufrêne, 2005). Les trames thématiques ont par conséquent, le rôle de représenter une série d'habitats d'un groupe d'espèces ayant des besoins écologiques similaires. L'avantage est qu'il doit très probablement exister une ou plusieurs espèces indicatrices pour chacune de ces trames. Celles-ci permettraient à leur tour de concentrer les efforts de conservation sur un plus petit nombre d'espèces si elles respectent bien les critères d'espèce parapluie. Une méthode particulière sera présentée dans ce travail afin de les déterminer et de construire un réseau écologique.

Le milieu d'étude est l'entièreté de la région Wallonne qui est la partie sud de la Belgique. Celle-ci s'étend sur une surface de 16 901 km<sup>2</sup> et possède une population de 3 633 795 habitants donnant une densité de 215 habitants au km<sup>2</sup> (Iweps.be).

Ce travail se concentre pour sa part, uniquement sur le réseau des oiseaux. Ces derniers sont pour la majorité, protégés en Belgique (Parlement Belge, 2019) et en Europe (Parlement Européen, 2009). Ils furent parmi les premières espèces à acquérir un statut de protection grâce à cette directive. De plus, de nombreux travaux les qualifieraient d'espèces parapluies ou en utiliseraient une comme exemple servant à la création d'un réseau écologique (Jongman et al., 2011; Duflot et al., 2018). Des questions pourraient toutefois se poser quant à la pertinence de l'élaboration d'un réseau écologique pour ce genre d'espèces. En effet, elles sont capables de parcourir de très grandes distances sans demander d'éléments particuliers pour se déplacer. Or les corridors font partie intégrante d'un réseau écologique et permettent de maintenir les liens vitaux au sein d'une métapopulation (Saunders et al., 1991). Il a déjà été prouvé qu'ils permettent au moins d'améliorer et de guider les mouvements des oiseaux disperseurs de graines (Tewksbury et al., 2002; Levey et al., 2005). De ce fait, un réseau centré sur les oiseaux aurait tout de même du sens. Effectivement, même s'ils ont beaucoup plus facile à se déplacer au sein de cette matrice que leurs homologues terrestres, ils ont tout de même besoin de se reposer, se nourrir et se reproduire. Ces processus biologiques se déroulent théoriquement à l'intérieur des zones cœurs. Un problème pourrait advenir si elles sont trop éloignées les unes des autres. Ceci est cependant, rarement le cas en Belgique car un haut degré d'anthropisation y est présent depuis des siècles et les espèces toujours présentes s'y sont relativement bien adaptées. Le pic noir est un bon exemple car il s'est habitué à occuper des forêts plus petites et assez éloignées les unes des autres (Jacob et al., 2010). A contrario, le pic épeichette serait plus sensible à la fragmentation des forêts feuillues. En restant dans un point de vue global, peu d'espèces d'oiseau sont dépendantes de couloirs naturels et donc le maintien de corridors pour faciliter leur déplacement ne serait à première vue, pas nécessaire. Il semblerait que l'importance de ces corridors pour eux soit bien plus grande au sein de paysages « récemment fragmentés » comme en forêts tropicales (Sieving et al., 2000) ou autres grandes forêts initialement homogènes (Rich et al., 1999).

Par ailleurs, les demandes écologiques de l'avifaune belge semblent assez diversifiées ce qui pourrait compliquer l'utilisation de trames thématiques (Jacob et al., 2010). Malgré tout, ce principe ne doit en aucun cas représenter une contrainte à l'élaboration du réseau écologique mais bien un outil d'aide à sa réalisation.

Les populations d'oiseaux nicheurs en Wallonie sont tout de même globalement en nette diminution et particulièrement pour les oiseaux agricoles (Derouaux & Paquet, 2018). De plus, le changement climatique actuel reste un élément important participant également à cette diminution tout comme l'augmentation de l'anthropisation du paysage (Laudelout & Paquet, 2014). Ces points ne sont que quelques exemples parmi d'autres qui démontrent la nécessité d'agir pour la conservation de l'avifaune. Ainsi, même si l'importance des corridors dans le contexte wallon est encore à prouver, la définition des zones cœurs de biodiversité reste une nécessité. Ce travail permettrait en plus de réviser les différentes mesures de restauration déjà mises en place mais de leur donner plus de poids ou de les corriger (Dufrêne, 2005).

## 1.1. *Acquisition et prétraitement des données*

L'objectif majeur du réseau écologique à construire est qu'il contienne les zones principales de nidification du plus grand nombre d'espèces d'oiseaux nichant en Wallonie. Pour se faire, deux sources différentes de données sont utilisées. La première regroupe deux plateformes de recensement publiques offrant des observations dites courantes (Paquet et al., 2013). Il y a la plateforme naturaliste de Belgique nommée "Observations.be" gérée par l'ASBL Natagora, Natuurpunt et la fondation néerlandaise (observation.be, 2020) et celle proposée par le DEMNA qui est l'OFSH. Elles permettent toutes deux aux observateurs lambda d'encoder les espèces qu'ils viennent d'observer en précisant le lieu, la date et diverses informations complémentaires. Ces données sont ensuite centralisées et la majorité est vérifiée par des experts. Ces données scientifiques offertes par les citoyens se sont déjà montrées fortement utiles dans la conception de zones de conservation (Elith et al., 2009). Cependant, étant donné que le nombre d'individus n'est pas systématiquement référencé, il est nettement préférable de considérer la nature de ces données et de les traiter comme étant de présence uniquement. La deuxième source provient de l'Atlas européen des oiseaux nicheurs en Wallonie. Celui-ci est également géré par l'AVES (pôle ornithologique de Natagora) et consiste à un recensement exhaustif des oiseaux au sein de carrés kilométriques dispersés en Wallonie (Aves.be, 2015). Ces deux sources sont fondamentalement différentes car celle-ci offre des données d'abondance. En effet, malgré la grande quantité de données disponibles, les observations courantes n'informent généralement pas sur le nombre d'individus mais seulement sur sa présence. Elles ne représentent également pas les mêmes espèces. En effet, la liste d'espèces nicheuses en Wallonie peut être séparée en deux groupes. Le premier est constitué des espèces les plus abondantes en Belgique (Annexe 2) dont les données relatives proviennent de l'Atlas. Le second regroupe les espèces jugées rares (Annexe 1) dont leurs données proviennent des plateformes de recensement. Ces espèces ne sont pas assez représentées dans les inventaires réalisés pour l'Atlas, c'est pourquoi il est préférable d'utiliser les données d'observation civiles.

Les données courantes ont été scrupuleusement prétraitées et triées selon plusieurs critères par le personnel de Natagora telles que la certitude de la donnée, sa précision et sa date d'observation. Cette dernière permet d'identifier les observations ayant été prises durant la période de migration de l'espèce et de les mettre à l'écart. Ceci permet d'éviter que des individus uniquement de passage en Wallonie ne faussent les résultats.

## 1.2. *Modélisation*

Toujours concernant uniquement les données des espèces « rares », même si elles sont d'une quantité assez considérable, elles ne couvrent pas l'entièreté de la Wallonie et ne proviennent pas d'échantillonnages rigoureux. De ce fait, l'identification des espèces indicatrices ainsi que la délimitation d'un réseau écologique via celles-ci ne semblent pas adéquates sans une couverture du territoire plus exhaustive. La stratégie proposée repose sur la construction d'un modèle d'estimation de la distribution pour chaque espèce afin d'estimer leurs habitats favorables ainsi que leur probabilité de présence sur l'entièreté de la zone d'étude.

Ces modèles de distribution d'espèce tentent de représenter la niche écologique fondamentale d'une espèce en particulier (Pearson et al., 2004). Cette dernière regroupe les différents milieux favorables à l'espèce et regroupe ainsi les différents habitats où elles pourraient s'y trouver (Hutchinson, 1957). Dans la plupart des cas cependant, les modèles ne parviennent qu'à estimer la niche potentielle de l'espèce (Peterson et al., 2001).

Cette niche est définie comme étant la combinaison spatio-temporelle possible des facteurs environnementaux déterminant la niche fondamentale. En effet, cette dernière regroupe l'entière des différentes combinaisons de facteurs environnementaux permettant le maintien d'une espèce à un endroit donné. Cependant, toutes ces combinaisons ne sont pas forcément possibles à un certain lieu à un certain moment d'où la définition de ce second type de niche (Jackson et al., 2000). La niche réalisée quant à elle, de par sa définition, est bien plus complexe à estimer. En effet, elle prend en compte les relations interspécifiques et l'aspect historique des milieux (Wiens, 1989) qui sont des variables difficiles à cartographier. Ceci signifie par la même occasion que la niche fondamentale peut détenir des zones où l'espèce y a été exclue par compétition, qu'elle y est éteinte ou qu'elle n'est pas parvenue à s'y disperser (Anderson et al., 2003). De plus, la niche réalisée peut comporter des milieux qui ne sont pas les plus optimaux pour l'espèce mais qui sont tout de même occupés dus à une exclusion au sein de sa niche fondamentale. Il y a donc des différences notoires entre la niche réalisée et la potentielle.

La modélisation implique trois grandes hypothèses. La première est que la relation entre la distribution de l'espèce à modéliser et les variables qui l'expliquent sont à l'équilibre (Guisan & Edwards Jr, 2002; Guisan et al., 2017). En effet, si l'espèce est toujours en expansion géographique, cela peut amener à des biais d'estimation. La seconde hypothèse est que toutes les variables influençant cette distribution sont prises en compte au sein du modèle. L'absence d'un prédicteur important peut amener à une variance très forte (Austin et al., 2011). La dernière est la pertinence des observations par rapport à l'objectif du modèle. Cela veut dire par exemple, que le stade de développement de l'individu observé doit être le même que celui devant être modélisé (Guisan et al., 2017). D'autres hypothèses doivent être posées et celles impactant le plus le cas d'étude seront discutées plus loin.

Deux différents algorithmes seront testés pour la modélisation des espèces rares. Il s'agit de MaxEnt (Maximum Entropy) qui fonctionne très bien avec des données de présence uniquement (Phillips et al., 2006) et de GAM (modèle additif généralisé). Pour cette dernière, une sélection améliorée des absences est nécessaire car elle demande de vraies absences en données d'entrée (Smolik et al., 2010).

### 1.2.1. Maxent

La méthode de MaxEnt consiste à estimer une densité de probabilités de présence en trouvant celle qui maximise l'entropie dans un espace géographique (Phillips et al., 2006). Par conséquent, l'hypothèse la plus simple sera toujours considérée comme étant la bonne et qu'en cas d'informations manquantes, l'événement est équiprobable. Autrement dit, le but de MaxEnt est de minimiser l'entropie relative entre deux densités de probabilités. Ces dernières correspondent aux présences et aux variables environnementales (Elith et al., 2011). Quoi qu'il en soit, il s'agit d'un algorithme de Machine learning qui s'alimentera de toutes les informations données afin de s'améliorer tout en respectant les contraintes qui lui sont imposées. Celles-ci veillent à ce que la moyenne empirique d'une variable environnementale soit la plus proche possible de la moyenne après modélisation. Par exemple, l'altitude moyenne des points d'observation doit être la plus proche de celle mesurée sur les points de présence estimés (Merow et al., 2013). Plus concrètement, l'algorithme cherche à comprendre la façon dont les distributions connues des espèces répondent aux facteurs environnementaux afin d'estimer le reste de ces distributions qui est inconnu (Elith et al., 2011). Il compare donc les zones de présences connues à des zones d'absences supposées appelées pseudo-absences qui composent le background de l'algorithme (Warren et al., 2014).

Pour cela, ces variables vont subir par modèle, une transformation mathématique qui expliquera au mieux la présence de l'espèce (Merow et al., 2013). Ces transformations sont des types de fonctions bien particuliers et il y en a six; linéaire, produit, quadratique, charnière, seuil et catégorique (Elith et al., 2011). Les résultats offerts par cet algorithme représentent un degré d'appartenance du pixel à une des deux catégories possibles qui sont l'absence ou la présence (Lobo et al., 2008).

### 1.2.2. GLM et GAM

La modélisation linéaire généralisée est une analyse de la régression entre une variable à expliquer et une ou plusieurs variables explicatives. Elle permet au contraire des régressions linéaires, d'être utilisée lorsque l'hypothèse de linéarité n'est pas respectée (Harezlak et al., 2018). Ainsi, elle n'oblige pas la transformation des données à une échelle non naturelle, car le lien entre le prédicteur et la distribution ainsi que la distribution elle-même ne doit pas être exclusivement linéaire (Hastie et al., 1990). Ces modèles se basent sur une relation hypothétique entre la moyenne de la variable expliquée et une combinaison linéaire des variables explicatives (Guisan et al., 2002). Ils se basent donc sur des modèles linéaires mais assouplissent l'hypothèse première de ces derniers où la moyenne conditionnelle de la réponse doit être linéaire par rapport à la variable explicative (Guisan et al., 2002; Harezlak et al., 2018). Les variables explicatives ou prédictives sont combinées entre elles afin de produire une variable explicative linéaire. Celle-ci est liée à la valeur attendue de la variable expliquée par une fonction particulière. Grâce à tout cela, d'autres types de distribution que la gaussienne peuvent être ajustées comme celles de la famille de poisson ou des binomiales. Cette dernière concerne justement la modélisation de distribution d'espèce dans laquelle 0 équivaut à une absence et 1 à une présence. Elle se présente sous la forme suivante (Harezlak et al., 2018) :

$$\log\left(\frac{p}{1-p}\right) = \beta + f(X_1) + \dots + f(X_n)$$

La méthode de modélisation additive généralisée est une extension semi-paramétrique de la régression linéaire générale (Guisan et al., 2002; Zaniwski et al., 2002; Santika et al., 2009). Elle peut aussi être considérée comme un mixte entre un modèle général linéaire et un modèle additif (Harezlak et al., 2018). Les fonctions similaires à la méthode précédente sont comme le nom l'indique, additives (s'additionnent entre elles) et lissées. Par conséquent, cette méthode ajuste indépendamment les variables prédictives via des fonctions lissées plutôt que de supposer une relation linéaire ou quadratique entre elles et la réponse (Zaniwski et al., 2002). Cette méthode suit donc le même principe que la précédente en déterminant une relation entre la moyenne de la variable expliquée et une fonction lissée des variables explicatives. Elle est forte utile lorsque des réponses entre les variables prédictives et expliquées ne sont justement pas linéaires. De fait, sa force principale est de gérer des relations hautement non linéaires ou non monotones entre la variable expliquée (présence/absence) et l'ensemble des variables (Guisan et al., 2002). Elle est capable d'identifier directement les transformations adéquates à réaliser sur les variables prédictives et donc de prévoir des réponses non linéaires afin de caractériser au mieux le lien entre les différentes variables environnementales et la présence de l'espèce. Ainsi, le logit de la formule du GLM n'est plus linéaire mais représente une fonction additive des variables explicatives. Généralement, cette méthode offre de meilleurs résultats en matière de prédiction (Santika & Hutchinson, 2009). Le terme semi-paramétrique apparaît car certaines variables peuvent être imbriquées dans le modèle de façon non-paramétrique. Par contre, GLM et GAM nécessitent tous deux des données de présence et absence (Zaniwski et al., 2002).

### 1.2.3. Sélection des variables

Les données environnementales qui vont servir de variables explicatives lors des modélisations proviennent de la couche «écotopes» (Radoux et al., 2019) qui ne compte pas moins de 106 variables. Celles-ci vont des variables bioclimatiques aux variables d'occupations du sol en passant par la texture des sols. Initialement, la couche des écotopes représente des polygones dont les caractéristiques environnementales sont homogènes et où les occupations du sol sont exprimées en proportion. Parmi toutes ces variables, un tri doit être fait car lorsqu'un trop grand nombre d'entre elles est pris en compte, les modèles perdent en efficacité (Zeng et al., 2016). Il est donc nécessaire d'y sélectionner une série de variables qui regrouperait les plus explicatives de la présence de chaque espèce (Elith et al., 2011). Cependant, le nombre n'est pas le seul problème. La colinéarité et la corrélation entre les différentes variables environnementales doivent être supprimées (Dormann et al., 2013; Merow et al., 2013; Warren et al., 2014) ou au moins minimisées (Guisan et al., 2002; Elith et al., 2011; Palaoro et al., 2013). La présence de cette colinéarité signifie que les variables en question ne sont pas indépendantes l'une de l'autre. Cela peut soit s'expliquer par le fait qu'elles expriment toutes les deux le même processus naturel ou que l'échantillon est trop faible causant l'absence de certaines combinaisons de variables (Dormann et al., 2013). Des problèmes apparaissent lors de la postposition du modèle à un niveau où la corrélation n'est plus la même. Ceci peut apparaître lors de son utilisation dans une aire géographique différente ou en utilisant des conditions environnementales futures. Alors, il y a une exacerbation des estimations et une grande diminution de la qualité du modèle (Harrell, 2001; Dormann et al., 2013). Cette étude-ci travaille sur une zone relativement grande qui s'étend sur plus de 16 000 km<sup>2</sup>. De ce fait, il y a une grande probabilité que les corrélations entre les variables varient en fonction du lieu. Par exemple, la température et la topographie sont, dans certaines situations géographiques, fortement liées mais dans d'autres, ça n'est plus le cas (Guisan et al., 2017). De plus, deux variables colinéaires pourraient ne pas avoir d'impacts significatifs au sein du modèle alors que mises séparément, le contraire apparaîtrait (Guisan et al., 2002). À l'opposé, une variable non significative fortement corrélée à une variable importante pourrait également être sélectionnée au sein d'un modèle (Dormann et al., 2013). Finalement, la paramétrisation des modèles peut fortement varier en changeant légèrement le jeu de données de base dû à ces corrélations. Ce qui est très dommageable (Dobson, 2002).

Tout ceci justifie l'utilisation de techniques spécifiques à la résolution de ces problèmes. De nombreuses méthodes ont d'ores et déjà été proposées afin de sélectionner efficacement les variables. Celles-ci permettraient d'une part de sélectionner celles qui participent le plus au modèle et d'une autre part, qui ne soient pas corrélées entre elles (Warren et al., 2014). Cette sélection peut se faire en suivant deux principes différents. Le premier consiste en une sélection basée sur les connaissances de l'écologie, la biogéographie et des autres éléments capables d'influencer la distribution de l'espèce. La seconde consiste quant à elle à se baser sur des processus itératifs de sélection via différents critères statistiques. La plupart préconisent pour se faire, l'utilisation d'analyse de la composante principale (PCA) sur un jeu de variables restreint afin de créer des groupes de variables liées entre elles (Merow et al., 2013). Il y a également possibilité de regarder directement la colinéarité des variables via une analyse de la corrélation (Merow et al., 2013; Warren et al., 2014). Ensuite, celles qui participent le plus au modèle sont maintenues. Très souvent aussi, la méthode du stepwise est utilisée mais montrerait une grande variabilité et serait donc à utiliser avec prudence (Guisan et al., 2002).



Pour finir, un point particulier doit être précisé concernant l'algorithme de MaxEnt. Comme il a été précisé précédemment, l'algorithme cherche à égaliser la moyenne estimée de chaque variable avec la moyenne empirique (Merow et al., 2013). Cependant, l'obtention d'une parfaite égalité a de grandes chances de mener à un surajustement de l'aire de distribution. C'est pourquoi chaque fonction utilisée au sein du modèle liant les variables explicatives à la variable à expliquer subit une régularisation. Ceci contraint la distribution modélisée à avoir une valeur moyenne de chaque variable au sein d'un certain intervalle autour de la moyenne empirique (Warren et al., 2011; Merow et al., 2013). Le software actuellement disponible de MaxEnt propose un paramètre de régularisation par défaut pour chaque type de fonction cité plus haut. Ceux-ci ont été calibrés sur un jeu de données assez conséquent afin de prendre en compte le plus de situations possible (zones géographiques et groupes spécifiques divers) (Philips et al., 2008). Il est cependant possible de les modifier via un facteur multiplicateur de ces paramètres qui se nomme le paramètre  $\beta$  (Warren et al., 2011). Ces paramètres par défaut ne font guère l'unanimité dans la littérature et certains préconiseraient une sélection pointilleuse de  $\beta$  selon les objectifs de l'étude (Warren et al., 2011, 2014; Merow et al., 2013; Zeng et al., 2016). Cette régularisation a un impact direct sur la complexité d'un modèle car au plus celle-ci est élevée et au plus des fonctions seront enlevées du modèle avec l'effet négatif d'éloigner la moyenne empirique de celle estimée (Merow et al., 2013). Ainsi, en modifiant ce paramètre, il est possible de contrôler la complexité du modèle. Ceci est très intéressant dans la mesure où un modèle trop complexe a tendance à surajuster les données, ce qui le rend trop spécifique aux conditions dans lesquelles il a été construit (Warren et al., 2014). Il serait donc préférable de modifier  $\beta$  afin qu'il soit plus adapté aux conditions de l'étude. Malgré tout, les paramètres par défaut ont montré leur efficacité à maintes reprises et se basent sur des fondements solides (Philips et al., 2008). De plus, il serait bien plus efficace de travailler avec ce dernier afin d'économiser du temps et de plus se concentrer sur la sélection des variables et du background. En effet, il ne semble pas si dommageable d'agir ainsi. L'objectif étant d'estimer la distribution d'espèces au sein d'une aire d'étude bien définie, l'obtention de modèles trop spécifiques à la zone n'est pas problématique (Merow et al., 2013). Également, il semblerait que la sous-paramétrisation d'un modèle soit nettement plus problématique que l'inverse qui se limite généralement à une potentielle sous-estimation de la distribution d'une espèce. Ainsi, comme l'objectif final est de désigner des zones de protection, il serait préférable de bien choisir les variables à utiliser et de ne pas être restrictif dans la régularisation. Cela pourrait résulter sur l'obtention de niches écologiques potentielles trop généreuses (Warren et al., 2014).

#### 1.2.4. Biais d'échantillonnage

Un modèle d'estimation de la distribution spatiale d'une espèce permet d'estimer l'adéquation d'une espèce en fonction de la valeur de certaines variables environnementales (Phillips et al., 2006). Par conséquent, avec un échantillonnage des présences (et absences selon la méthode de modélisation utilisée), il est possible d'estimer la distribution de l'espèce sur l'entièreté de la zone géographique étudiée. Cependant, les données des espèces rares proviennent d'observations courantes et donc d'échantillonnage non rigoureux. Par conséquent, l'absence d'une espèce à un endroit peut s'expliquer de deux façons différentes. Elle est soit due à l'environnement qui n'est pas propice à l'espèce ou bien à un effort d'échantillonnage trop faible. Les présences se justifient par les phénomènes inverses. Ainsi la probabilité d'occurrence d'une espèce se décompose en trois composantes, la probabilité que le site soit inventorié, la probabilité que l'espèce soit détectée et la probabilité que l'espèce soit présente à ce moment-là (Merow et al., 2013). Comme les deux premières probabilités ne sont pas constantes en fonction des sites, un biais d'échantillonnage non négligeable apparaît systématiquement au sein des données (Yackulic et al., 2013).

Or, elles sont censées ne pas être biaisées (Guisan et al., 2017). Ce biais peut être séparé en deux catégories distinctes. Il y a dans un premier temps, les biais géographiques dus aux contraintes influençant la réalisation ou non d'un échantillonnage à un endroit donné. Il représente les zones ayant reçu un plus grand effort d'échantillonnage dû à leur situation géographique plus favorable, telle que leur facilité d'accès, leur réputation, etc. La seconde regroupe les biais environnementaux apparaissant lorsque tout l'éventail de variations environnementales n'est pas pris en compte car chaque type d'habitat n'est pas forcément visité (Barbet-Massin et al., 2012). Cependant, cette dernière composante est assez limitée en vue de la grande quantité de données offertes par Natagora sans pour autant la supprimer totalement. En effet, travailler sur une grande aire géographique et avec un grand nombre de données permet de mieux prendre en compte l'amplitude écologique d'une espèce et ne pas sous-estimer ses habitats favorables potentiels. De ce fait, la principale cause de biais provient de cet effort d'échantillonnage inconstant. Ceci ne permet donc pas d'assurer l'absence d'une espèce pouvant générer des fausses absences. Également, de fausses présences peuvent s'être insérées car la vérification des observations par des experts n'est premièrement pas systématiques et deuxièmement, peut s'avérer imparfaite. De plus, certaines variables environnementales ont pu directement influencer la détectabilité de certaines espèces et renforcer le biais. En effet, une variable positivement colinéaire avec la probabilité de détectabilité provoquerait une surestimation de la distribution de l'espèce alors qu'une relation négative provoquerait l'inverse (Yackulic et al., 2013). Il s'agit pour finir, d'espèces à forte mobilité réduisant ainsi leur détectabilité et la précision des observations (Barbet-Massin et al., 2012). Ceci met notamment en avant l'importance de travailler avec des absences et présences ayant un biais similaire. En effet, cela permettrait de réduire les effets négatifs du biais (Elith et al., 2011). Fort heureusement, quelques articles traitent du sujet et ont tenté d'apporter des solutions à ce problème de biais tels que Fourcade et al., (2014) qui présentent cinq méthodes afin de contrer celui-ci. Pour finir, de nombreux autres articles démontrent que la minimisation du biais d'échantillonnage se fait principalement durant la sélection des pseudo-absences insérées au sein du background pour MaxEnt (Barbet-Massin et al., 2012; Merow et al., 2013; Yackulic et al., 2013; Fourcade et al., 2014). Pour pouvoir utiliser les deux autres méthodes de modélisation, il en va de soi que la véracité des absences sélectionnées doit être forte.

### 1.2.5. Sélection du Background (pseudo-absence)

La sélection des pseudo-absences est une étape cruciale dans l'obtention d'un modèle de qualité. D'autant plus qu'un algorithme de modélisation nécessitant de vraies absences est testé ce qui demande une attention encore plus particulière à cette sélection. Même si GAM a déjà été utilisé avec des données de présence uniquement lorsque MaxEnt n'était pas encore très développé, les estimations générées furent souvent médiocres (Zaniewski et al., 2002).

Une fois de plus, la littérature est assez fournie en méthodes diverses afin de sélectionner ce background mais aucune ne semble faire l'unanimité. Tout d'abord, certains préconisent l'utilisation de données de présence d'autres espèces du même groupe qui auraient l'avantage de facilement être accessibles dans ce travail (Yates et al., 2010; Merow et al., 2013; Warren et al., 2014). Ceci permettrait d'égaliser le biais d'échantillonnage des présences à celui des pseudo-absences (Phillips et al., 2009; Elith et al., 2011). Cependant, il faut émettre l'hypothèse forte que l'espèce étudiée aurait été recensée si elle avait été vue par l'observateur à l'endroit où l'autre espèce a été observée (Merow et al., 2013). De plus, cette méthode risque fortement de mettre plus en évidence la différence de zones géographiques occupées par deux espèces différentes que celle entre un habitat favorable et défavorable.

Ce défaut est d'autant plus vrai que si la distribution de ces deux espèces suit une dynamique de compétition/exclusion plutôt qu'une dynamique d'adéquation à l'habitat (Warren et al., 2014).

De plus, cette technique communément appelée « Target group » donne un background avec des points spatialement autocorrélés. Les points de validations ne sont donc plus indépendants l'un de l'autre (Elith et al., 2009). Ceci cause une augmentation non négligeable de la capacité discriminante apparente du modèle sans pour autant améliorer sa capacité prédictive (Veloz, 2009; Gonzalez et al., 2011).

Finalement, d'autres conseilleraient la sélection de 10 000 points qui est le nombre par défaut dans MaxEnt (Phillips et al., 2006). Or il n'est également pas accepté par tout le monde. Il semblerait plus judicieux d'utiliser un nombre de pseudo-absences équivalent au nombre de présences disponibles afin d'égaliser le plus possible leur biais d'échantillonnage respectif et d'ainsi avoir une prévalence de 0,5 (Hengl et al., 2009). Cette prévalence se définit comme la proportion de présences au sein de l'échantillon (Lawson et al., 2014). Un trop gros background pourrait augmenter la proportion de fausses absences provoquant une sous-estimation de sa niche écologique potentielle (Anderson et al., 2010). Afin de réduire cette proportion, il est plus avisé de sélectionner ces points en dehors d'un tampon construit autour des points de présence (Mateo et al., 2010). À l'opposé, réduire drastiquement le nombre de points d'absence pourrait provoquer une surestimation de l'adéquation de l'espèce aux zones non échantillonnées dont le nombre augmente indubitablement (Iturbide et al., 2015).

En raison du manque de consensus scientifique à ce sujet, plusieurs méthodes de sélection du background seront testées et une analyse de sensibilité effectuée. La première proposée consiste à utiliser des pseudo-absences choisies aléatoirement sur toute l'aire d'étude et d'un nombre total de 10 000 points (Phillips et al., 2006). Celle-ci sert de point de repère et permettra d'évaluer la qualité du background amélioré lors de son utilisation avec MaxEnt mais surtout avec GAM qui est nettement plus sensible aux absences. Ainsi, les deux méthodes suivantes ont comme objectif dans un premier temps de diminuer le biais d'échantillonnage au sein des observations et dans un second temps, d'améliorer les absences afin de pouvoir les considérer comme de « vraies absences ». Celles-ci consistent à utiliser deux critères distincts comme proxy de l'effort d'échantillonnage du site. La première se base sur le nombre d'espèces différentes recensées au sein du même site. Si cette diversité spécifique dépasse un certain nombre, alors le point central de ce site est éligible comme absence. Le proxy de la seconde méthode est le nombre d'observations par site. Ces deux méthodes se basent sur la même hypothèse forte que le « Target group » (Iturbide et al., 2015) stipulant que l'espèce à modéliser aurait été recensée si elle avait été présente. Cette hypothèse s'assouplit légèrement car la modélisation dans le cadre de ce travail se fait sur des espèces rares. Par conséquent, il y a plus de chances que celles-ci soient recensées dès qu'elles sont vues. Elles sont également plus recherchées mais a contrario, ont plus tendance, pour certaines, à être cryptiques ce qui réduirait l'efficacité de ces deux méthodes. Dans cette optique, il y a deux façons de penser qui sont illustrées par ces deux proxys différents. La première serait de penser qu'un grand nombre d'espèces observées à un endroit augmenterait la probabilité d'un inventaire exhaustif. La seconde part sur le principe qu'au plus un lieu a été visité et au plus il y a des chances qu'une espèce n'y ayant pas été observée est due à l'environnement du site plutôt qu'à un effort d'échantillonnage insuffisant.

Pour terminer, comme le préconise la littérature, il est préférable d'utiliser le même nombre d'absences que de présences (Hengl et al., 2009). Cependant, en raison du nombre de présences assez limité pour de nombreuses espèces, les modèles résultants ne sont guères envisageables et cette façon de procéder est écartée. En effet, une distribution des absences trop contraintes amène souvent à des modèles biaisés (Vanderwal et al., 2008).

### 1.2.6. Évaluation des modèles

La dernière étape de la modélisation consiste à évaluer le modèle résultant. Au fil des années, divers indices et critères ont été développés afin de parvenir à évaluer au mieux un modèle donné. La méthode la plus simple est l'élaboration d'une matrice de confusion qui permet de mettre en avant le nombre d'éléments ayant été correctement prédits. Cependant, cette matrice nécessite que chaque donnée appartienne à une classe bien définie (présence et absence). Pour cela, il faut déterminer un certain seuil permettant de classer les entités résultantes de l'algorithme. Or, dans la plupart des cas lors de la détermination de l'aire de distribution d'une espèce, le nombre de points d'absence n'est pas égal au nombre de points de présence. Cela empêche l'utilisation habituelle d'un seuil de 0,5. Une prévalence ainsi différente affecte énormément l'évaluation du modèle via une simple matrice de confusion et un seuil pris au hasard (Jiménez-Valverde et al., 2007; Liu et al., 2013). La meilleure alternative proposée par la littérature est l'utilisation et la construction de la ROC (receiver operating characteristics) grâce à un jeu de données de validation du modèle (Fawcett, 2006). Ceux-ci correspondent à des points d'absence et de présence différents et indépendants des points ayant servi à la construction du modèle et ayant la même prévalence. La courbe, quant à elle, représente l'évolution du taux de vrais positifs en fonction du taux de faux positifs selon les différents seuils de différenciation entre une valeur positive et négative. Cette capacité représente un très grand avantage car elle prend en compte tous les différents seuils possibles allant de 0 à 1 (Fielding et al., 1997; Baker et al., 2001; Lobo et al., 2008; Chicco et al., 2020). La borne inférieure considère tous les points comme positifs et la borne supérieure, comme négatifs. Cette méthode a été particulièrement développée dans le cadre de la médecine afin de comparer la performance de tests cliniques (A.Hanley et al., 1982; A.Swets, 1988; Pepe, 2000) mais également pour des algorithmes de machine Learning (Spackman, 1989; Huang et al., 2005). En effet, il peut s'appliquer à tout modèle qui prédit une catégorie homogène au sein de chaque pixel d'une carte (Pontius et al., 2001). Ainsi, en visualisant la manière dont la courbe évolue, il est possible d'évaluer la performance d'un modèle ou d'un test clinique.

D'une façon plus générale, c'est l'AUC qui sert plus directement d'indice de performance d'un modèle de distribution d'une espèce et qui est utilisé le plus fréquemment dans ce domaine (A.Swets, 1988; Osborne et al., 2002; Lobo et al., 2008; Veloz, 2009; Gonzalez et al., 2011). Ce dernier est la mesure de l'aire sous la courbe ROC. Il est borné de 0 à 1 où la valeur de 1 signifie une précision parfaite alors qu'une valeur de 0,5 indique que le modèle a la même performance qu'un modèle aléatoire. Si cette valeur est inférieure à 0,5, cela signifie par contre qu'il estime les présences et les absences d'une moins bonne façon qu'un modèle aléatoire (Veloz, 2009). Il a été proposé pour la première fois dans le domaine de la modélisation d'aires de distribution d'espèces par Fielding & Bell (1997) et fut conseillé de nombreuses fois depuis.

Cependant, cet indice possède de nombreux défauts et se baser uniquement sur lui pour évaluer un modèle n'est pas conseillé (Austin, 2007). L'AUC démontre la capacité discriminante d'un modèle ce qui se traduit par la probabilité qu'un élément positif soit correctement discerné d'un élément négatif. Cette propriété qui reste très intéressante ne permet pas à l'AUC d'évaluer l'ajustement du modèle à la distribution réelle de l'espèce (Vaughan et al., 2005).

Par conséquent, un modèle pourrait posséder un bon AUC sans pour autant bien ajuster la distribution de l'espèce. À l'inverse, un seuil élevé pourrait causer un faible AUC même si le modèle ajuste bien la distribution. De plus, deux parties de l'AUC ne sont pas intéressantes dans le cas présent qui sont la partie extrême gauche et extrême droite. Celle-ci représente respectivement une sensibilité nulle (taux de vrais positifs) et une spécificité nulle (taux de vrais négatifs).

La partie la plus intéressante est celle où se trouve le seuil qui sera finalement choisi car les autres parties ne servent plus qu'à gonfler l'AUC (Baker et al., 2001). Pour finir, les erreurs de commissions et d'omissions sont pondérées de la même façon alors que leur importance dépend clairement de l'objectif même du modèle (Anderson et al., 2003; Peterson, 2006; Lobo et al., 2008). Également, en raison du biais d'échantillonnage stipulé plus haut, les probabilités de fausses absences sont intrinsèquement plus élevées que celles des fausses présences ce qui modifie la ROC (Lobo et al., 2008). Pour ces différentes raisons, d'autres indices d'évaluation seront également pris en compte dans le cadre de ce travail.

### 1.2.7. Seuil de probabilité

Malgré tout, l'utilisation de l'AUC reste un atout considérable spécialement pour la désignation du seuil final permettant de convertir les données quantitatives offertes par MaxEnt en données binaires. Pour ce point, la littérature semble assez d'accord pour stipuler qu'il faut sélectionner le seuil qui maximise la sensibilité et la sensibilité au sein de la ROC (Jiménez-Valverde et al., 2007; Liu et al., 2013).

### 1.2.8. Autres critères d'évaluation

Une fois que le problème de sélection de seuils est franchi, il est à nouveau possible d'utiliser des indices se basant sur une matrice de confusion. Cependant, il y a toujours le souci principal provenant d'une proportion de présence inférieure à 0,5 qu'il faut prendre en compte. Ceci empêche l'utilisation de certains indices et d'autres seraient plus conseillés. Trois indices assez intéressants ont été sélectionnés afin de répondre à cette situation qui sont le « Producer Accuracy », « User Accuracy » et l'indice de Matthew. Le premier, également appelé « recall » (Fawcett, 2006) est le rapport entre le nombre d'entités d'une certaine classe classifiée correctement sur le nombre total d'entités de cette classe (Story et al., 1986). Il s'agit donc de la mesure de l'erreur d'omissions ou le taux de fausses absences. Le second correspond au rapport entre le nombre d'entités d'une certaine classe correctement catégorisées sur le nombre d'entités ayant été catégorisées dans cette classe. En d'autres mots, cela mesure l'erreur de commission ou le taux de fausses présences. Le dernier indice retenu est le coefficient de corrélation de Matthew. Celui-ci remplace la mesure conventionnelle de la précision d'un modèle habituellement mesuré avec l'indice d'« Accuracy » mais qui n'est pas adapté à une prévalence différente de 0.5 (Blagus et al., 2010). Ce coefficient l'est bien plus car il prend en compte la différence de nombre de points de présence et de points d'absence (Powers, 2011; Chicco et al., 2020). Son utilité est d'offrir une bonne appréciation de la précision globale d'un modèle en faisant un ratio entre ce qui a été correctement prédit sur l'entièreté des entités de validation. Pour finir, la complexité de chaque modèle est également prise en compte même si son importance apparente a quelque peu été réduite en raison de l'objectif de l'étude. En effet, un modèle trop complexe ne semble pas être un problème lorsque le modèle est uniquement utilisé dans les mêmes conditions que dans lesquelles il a été construit (Warren et al., 2014).

Généralement, la complexité d'un modèle est mesurée via son AIC (Akaike, 1973) qui est un indice pénalisant la surparamétrisation (Merow et al., 2013). Cependant, l'indice pouvant être utilisé avec un modèle construit via MaxEnt est l'AICc (Muscarella et al., 2014). Le principe de comparaison entre deux modèles est similaire à celui de l'AIC où la valeur la plus faible indique le modèle le moins complexe (Warren et al., 2011).

### 1.3. *Identification des espèces indicatrices*

La modélisation décrite ci-dessus permet d'identifier la grande majorité des habitats favorables de chaque espèce en palliant aux problèmes d'échantillonnage. Concernant les espèces communes, un travail précédent celui-ci s'est chargé de l'élaboration de modèle solide offrant des données d'abondances. Il s'agit maintenant de trouver les espèces pouvant être sélectionnées comme espèces indicatrices de trames afin de construire un réseau écologique sur base de sa distribution géographique. Généralement, l'indice Indval (Dufrêne et al., 1997) est utilisé dans ce but, principalement en raison de sa facilité d'utilisation et son efficacité. Cependant, il demande des sites ou des habitats précisément différenciés ce qui n'est point le cas ici. En effet, la seule manière de matérialiser les trames serait via les proportions d'occupation du sol. Ainsi, un carré kilométrique serait considéré comme appartenant à la trame possédant la plus grande proportion d'occupation du sol. Or cela ferait perdre énormément d'information et résumerait bien trop fortement l'occupation du sol au sein d'un carré kilométrique. C'est pourquoi il serait préférable d'analyser la corrélation entre la probabilité d'occurrence des espèces rares et l'abondance des espèces communes et la proportion des différentes occupations du sol caractérisant une trame. Pour cela, une analyse d'un gradient direct via une analyse de la redondance ou RDA serait le plus adéquate (Dufrêne et al., 2006; Borcard et al., 2008).

### 1.4. *Réseau écologique*

Une fois cette liste d'espèces obtenue, le réseau écologique sera déterminé en se basant sur la distribution de ces espèces cibles. Pour se faire, les habitats favorables de chaque espèce par trame seront désignés comme zone cœur. Il s'agit donc des zones où au moins une des espèces indicatrices est présente.

De nombreuses méthodes se sont développées au cours des dernières années pour l'élaboration d'un réseau écologique (Amsallem et al., 2013). Comme le but principal de celui-ci est de maintenir la connexion au sein d'une métapopulation, le plus idéal serait de travailler directement avec les données d'observation empiriques des mouvements de l'espèce étudiée. Malheureusement, ce genre de données est très rarement disponible (Hilty et al., 2004) et ne l'est également pas ici. Il faut donc se baser sur les différents modèles construits précédemment afin d'estimer les meilleurs corridors. La LCP (Least Cost Path) est l'une des méthodes les plus utilisées (Sawyer et al., 2011). Elle consiste à déterminer le chemin reliant deux zones cœurs qui représente le plus faible coût de déplacement pour l'espèce étudiée. Ainsi, une carte représentant les coûts aux déplacements d'une espèce au sein de la matrice paysagère est déterminée. Ces coûts peuvent être déterminés en se basant sur l'avis d'expert et les connaissances écologiques acquises sur l'espèce en question (Ayram et al., 2014). Cependant, il ne serait pas conseillé d'agir ainsi car certaines informations et processus clés pourraient ne pas être pris en compte. En effet, Bourdouxhe et al., (2019) soutiennent l'utilisation directe des données de présence pour la détermination des coûts de déplacement.

Cela offrirait des résultats plus représentatifs des chemins que pourraient effectivement emprunter les individus concernés. Il faudrait donc plus préférentiellement se baser sur les niches écologiques potentielles modélisées. Cependant, cette méthode ne prend pas suffisamment l'aspect paysager et la connectivité entre les taches. C'est pourquoi Duflot et al., (2018) proposent de fusionner cette méthode avec celle des graphes spatiaux. Cette dernière consiste à considérer tous les éléments du paysage comme étant des points, des lignes ou la matrice où ils sont inculqués (Bunn et al., 2000). Les points sont donc les zones cœurs de biodiversité et les lignes sont les corridors les reliant entre eux. En faisant ainsi, le défaut cité précédemment est compensé.

Cette fusion des modèles et de la théorie des graphes rejoint le travail réalisé par Bourdouxhe et al., (2019) où ils ont également utilisé le modèle de distribution de l'espèce afin de déterminer la carte de résistance de celle-ci. Cependant, en suivant ce procédé, il y a tout de même des risques de mal identifier les milieux favorables ou non aux mouvements d'une espèce (Chetkiewicz et al., 2006). De plus, la résolution de cette étude risque d'être trop grossière pour qu'une analyse des coûts aux déplacements soit bien utile. Les corridors qui seront identifiés auront une taille équivalente à la résolution du travail qui pourra être affinée par la suite dans d'autres travaux. Si celle-ci est trop grosse, il y a une plus grande probabilité que l'analyse ne parvienne pas à identifier correctement les chemins les plus optimaux. En effet, des éléments du paysage entre deux zones cœurs pourraient s'avérer suffisants pour permettre un bon flux entre eux mais n'avoir pas assez de poids pour que l'analyse les identifie. Il serait donc plus pratique, pour cette première étape du projet du moins, de se contenter d'une analyse spatiale et de déterminer les corridors via la plus petite distance euclidienne (Pascual-Hortal et al., 2007). Ceci peut être facilement réalisé à l'aide d'un logiciel développé dans le cadre de l'élaboration de la trame verte et bleue en France se nommant Graphab (Foltête et al., 2012).

### **1.5. Évaluation du réseau écologique**

Après la construction d'un réseau par trame thématique, il est nécessaire de l'évaluer. Selon le guide méthodologique (Dufrêne et al., 2019), il y a deux différents types de zones cœurs à distinguer, les zones existantes et potentielles. Celles appartenant au premier type peuvent être identifiées de trois manières différentes. Soit, elles sont d'ores et déjà sous protection légale comme Natura 2000 ou sous la loi de la conservation de la Nature, soit elles possèdent une forte biodiversité observée ou elles contiennent un ou plusieurs sites de grand intérêt pour la biodiversité. Les zones potentielles sont celles possédant une biodiversité exceptionnelle mais venant des données de présence modélisées.

Par la suite, le logiciel Graphab propose une série de métriques permettant une bonne évaluation et caractérisation d'un réseau écologique. Parmi celles-ci, trois en particulier ont été sélectionnés afin de caractériser chaque réseau obtenu à deux échelles différentes, la globale et locale. La première est décrite par deux indices, la Probabilité de connexion qui représente la probabilité que deux individus pris au hasard au sein du réseau puissent se rejoindre (Saura et al., 2007; Girardet et al., 2018) et l'indice d'Harary. Ce dernier mesure le rapport entre le nombre de liens et d'habitats présents dans un réseau. Sa valeur minimale apparaît donc lorsqu'il n'y a pas de liens et l'inverse a lieu lorsque le maximum de connexion est obtenu (Girardet et al., 2018). Ces indices pris ensemble permettent de bien évaluer la connectivité des réseaux via deux angles différents. À l'échelle locale, il y a le dPC qui est calculé pour chaque lien et tache. Il représente concrètement l'importance de chaque composante dans la connectivité du réseau et est calculé via un processus itératif de suppression où l'indice PC est mesuré en enlevant à chaque fois la composante pour laquelle l'indice est calculé. Il se divise en trois parties, dPCintra, dPCflux et dPCconnector.

Le premier informe la connectivité à l'intérieur des taches ce qui n'est point important dans le cadre de ce travail-ci car chaque habitat sélectionné est censé être majoritairement favorable aux espèces de la trame en question. Le second exprime la capacité de réservoir d'espèces d'une tache en ne prenant en compte uniquement les taches voisines. Il est une représentation du flux d'espèces et celui-ci est directement dépendant de la connectivité de la tache avec ses voisines mais également de la capacité d'accueil de celles-ci (URBAN et al., 2009). Dans le cadre de ce travail, cette capacité est assimilée à la superficie de la tache. La dernière quant à elle complète celle-ci en mesurant l'importance de l'élément par rapport à la connectivité totale du réseau. Seule cette dernière peut être mesurée concernant les corridors (Saura et al., 2010; Bourdouxhe et al., 2019).

Pour une évaluation plus directe des réseaux, deux critères de validation sont posés et doivent être respectés pour qu'un réseau puisse être considéré comme correct. Le premier critère stipule que les indices de diversité en oiseau au sein du réseau doivent être supérieurs à la valeur mesurée dans la matrice paysagère. Le second critère, quant à lui, voudrait que le réseau écologique représente fidèlement sa trame, c'est-à-dire, qu'il soit recouvert majoritairement par l'occupation du sol définissant sa trame.

Pour finir, la dernière étape du travail est de croiser le réseau écologique existant et potentiel afin de déterminer leur niveau de similitude. Comme expliqué précédemment, la proportion des terres wallonnes efficacement protégées est très faible et le réseau Natura 2000 qui s'étend sur une plus grande surface n'est pas suffisant pour contrebalancer. En comparant la distribution géographique de ces sites avec la délimitation des réseaux écologiques, il sera possible d'évaluer leur pertinence et la nécessité d'augmenter le nombre de sites protégés. Également, deux nombreux sites sont reconnus pour leur grand intérêt biologique de manière globale ou centré sur les oiseaux sans pour autant être protégés. En réalisant la même comparaison avec ces sites, il sera possible d'évaluer l'efficacité de leur désignation.





## II. Matériel et méthode

### 2.1. Acquisition des données d'observation

Les différentes manipulations et traitements statistiques ont tous été réalisés sur R studio version 3.6.3 ou sur le programme office Excel 2013. Concernant les géotraitements, la plupart ont été effectués sur Rstudio et le reste, via Arcmap version 10.7.1. Les données brutes des espèces dites « rares » dont la liste est disponible en annexe (Annexe1), proviennent d'observations faites par des citoyens dans un intervalle de temps variable selon l'espèce. Cette variabilité est due aux tris réalisés par Natagora afin d'offrir un jeu de données propres qui réfère uniquement aux individus nichant en Région Wallonne. Les données couvrent la période du 1<sup>er</sup> janvier 2013 au 31 décembre 2018. Elles centralisent principalement les observations enregistrées sur la plateforme du DEMNA (OFFH) et celles d'Observation.be. Cependant, les informations relatives à trois espèces proviennent de recensements normalisés mis en œuvre par le personnel de Natagora. Il s'agit des deux espèces du genre *Milvus* ainsi que le Tarier des prés. De plus, les données de présence des espèces communes (annexe 2) ont directement été fournies sous forme de modèles informant sur leur abondance au sein de chaque carré kilométrique de Wallonie. Les données brutes ne sont donc pas disponibles. Par conséquent, la totalité du jeu de données qu'elles soient sous forme de modèles ou d'observations directes, regroupe l'entièreté des espèces nichant en Wallonie. Les seules exceptions concernent les espèces invasives, côtières et domestiques qui ont été délibérément exclues. La Bernache du Canada, l'Ouette d'Égypte, la Perruche à collier, le Goéland cendré, la Mouette rieuse, la Sterne pierregarin et le Pigeon biset domestique ne sont donc pas pris en compte.

### 2.2. Gestion des données

Les espèces rares ont subi une série de filtres et de sélections avec des critères stricts afin de ne garder que les données sûres et relatives aux individus nichant effectivement en Wallonie. Ces filtres sont basés principalement sur la date des observations, leur qualité, leur précision géographique, leur vraisemblance et tout ceci, sur avis d'experts. Une seconde sélection a été faite afin d'enlever l'entièreté des données n'étant pas assez précise, c'est-à-dire, ayant une précision inférieure à la résolution de ce travail. Cette dernière est la même que celle utilisée pour la modélisation des espèces communes. Celle-ci est fixée à 1km<sup>2</sup> car les unités d'échantillonnage de l'Atlas européen de Wallonie sont des carrés kilométriques. À première vue, une résolution plus grossière allant à 3km<sup>2</sup> peut être suffisante pour une étude sur l'avifaune. Elle permettrait d'englober les différents milieux écologiques proches que l'espèce pourrait utiliser (Johnston et al., 2020). Or, la possibilité de travailler sur une échelle plus fine semble tout de même plus intéressante et permettrait de déterminer plus précisément les éléments influençant la distribution des espèces. Cependant, utiliser une échelle trop fine risquerait quant à elle, de prendre en compte exclusivement les habitats de nidification. Même s'il s'agit du but premier de ce travail, la protection d'une surface minimale d'aire de nourrissage est indispensable justifiant cette résolution finalement choisie qui s'ajuste bien à la situation.

Les données des deux espèces du genre *Milvus* ont été fournies sous forme de carré de surface de 40km<sup>2</sup> représentant les aires de nidification connues de l'espèce. Ainsi, tous les carrés kilométriques se trouvant à l'intérieur de ces zones sont interprétables comme des points de présence. Malheureusement, ceci ne permet pas une modélisation convenable de leur aire de distribution car l'échelle initiale est trop grossière.

Concernant le Tarier des prés, seuls les sites Natura2000 l'abritant ont été considérés comme des milieux favorables à l'espèce de façon plus ou moins certaine. Par conséquent, tous les carrés kilométriques interceptant ces sites sont interprétés comme des points de présence. Cependant, cela peut entraîner une surestimation du nombre de ces points. Contrairement aux espèces du genre *Milvus*, il a été possible de générer un modèle en raison de l'échelle bien plus fine de la délimitation des sites Natura 2000.

### 2.3. Analyse des données de présence brutes

Avant de passer à la modélisation des niches potentielles de chaque espèce, il est intéressant de réviser les trames écologiques présentées dans le guide méthodologique (Dufrêne et al., 2019). Pour se faire, trois étapes distinctes ont été menées à bien. Chaque espèce (rare ou commune) est classée via ces trames de la meilleure façon possible et de nouvelles sont proposées et ajoutées lorsque cela semble indispensable afin de bien caractériser les exigences écologiques de l'espèce. Certaines peuvent être caractérisées directement par des trames secondaires si cela semble pertinent et dans la quasi-totalité des cas, plusieurs trames doivent être utilisées pour ne caractériser qu'une espèce. Par conséquent, des trames hypothétiques sont testées et incorporées dans la première classification mais seules celles ayant une fréquence relative supérieure à 1% sont maintenues en tant que trame principale ou secondaire selon leur nature. Par exemple, la trame « milieux rocheux » est décrite comme principale même si seulement six espèces lui sont inféodées (supérieur au seuil de 1%) car elle ne pouvait être classée comme sous-classe d'une autre trame. Grâce à ce regroupement d'informations, les nouvelles et anciennes trames sont redéfinies et ceci de façon plus propre à l'avifaune. La troisième étape consiste à réitérer la première en y ajoutant les nouvelles trames afin de vérifier une seconde fois leur pertinence.

Chaque trame peut être associée à une ou plusieurs variables d'occupation du sol reprises des couches écotopes dont la liste est reprise dans le tableau 2.

Tableau 2: Listes des trames associées aux variables d'occupation du sol (écotope) les caractérisant via leur code ainsi que leur signification.

Trame principale	Trame(s) secondaire(s)	Variable(s) associée(s)
Bocages et lisières		COpen (couvert permanent de buissons et d'herbacée provenant d'adventice de forêts ou de jeunes plantations), Bocages (carte spécifique)
Eaux de surface		Water (surface d'eau permanente)
Forêts	Forêts feuillues	BroadLV (arbres décidues)
	Forêts résineuses	NeedlLV (arbres sempervirents)
Milieux agricoles	Prairies	MGrainin (surface de graminée monospécifique)
	Cultures	Plowed (sols arables)
Milieux anthropisés		Artif (surface artificielles et bâtiments)
Milieux ouverts	Milieux secs	Nopen (zone ouverte avec sol relativement sec),
	Milieux humides	WOpen (zone ouverte avec sol humide)
Milieux rocheux		Bare (sols nus généralement associés à des zones de carrière)

Ensuite, toutes les variables écotopes jugées importantes sont rastérisées à la résolution d'un kilomètre carré. Ceci est réalisé grâce à ArcGis 10.7.1 via une grille de référence représentant les carrées kilométriques de l'Atlas afin que chaque pixel des rasters résultants représente un des carrés de la grille. Cette correspondance est importante pour la complémentarité des modèles des deux groupes d'espèces.

## 2.4. *Modélisation*

Les deux différents algorithmes testés afin de modéliser les aires de distribution sont MaxEnt à l'aide de la fonction `maxent()` du package « `dismo` » et GAM avec la fonction et le package du même nom.

### 2.4.1. Sélection des variables

Cette sélection doit veiller à minimiser au maximum la corrélation et par conséquent, la colinéarité entre les différentes variables tout en maintenant celles qui sont les plus explicatives des présences. Pour cela, une présélection parmi les 61 variables disponibles au sein de la couche écotope redimensionnée à l'échelle d'un kilomètre carré (Radoux et al., 2019) a été réalisée sur base de leur pertinence écologique et sur la qualité de la variable elle-même. Ainsi, une série de variables est mise de côté car leur utilisation au sein d'un modèle « oiseaux » n'est pas écologiquement justifiable. C'est le cas par exemple de la variable « position » qui exprime l'altitude relative moyenne d'un écotope dans un rayon de 200m autour de lui. Il y a également les variables jugeant sur la qualité des variables écotopes qui sont à retirer. De plus, celles qui informent sur des distances sont également écartées car la résolution utilisée diminue fortement leur intérêt et ne les rend plus pertinentes.

De manière à analyser la corrélation entre les variables, la valeur des pixels de chaque variable est extraite du point central de chaque carré kilométrique. Grâce à cela, une matrice de corrélation entre les 56 variables restantes peut être générée. La sélection entre les variables corrélées est réalisée à l'aide d'une analyse de la composante principale (Borcard et al., 2008). Pour chaque ligne de la matrice de corrélation, les variables ayant une corrélation de Pearson (Benesty et al., 2009) supérieure à 70% (Green, 1979; Menard, 2002) sont regroupées afin de réaliser l'ACP. Celle participant le plus à l'axe principal est choisie et les autres sont écartées. Les coordonnées des vecteurs de chaque variable représentant leur projection sur l'axe principal informent sur la participation des variables et permettent cette sélection. Une attention particulière a également été portée à l'eigenvaleur (variance expliquée) de l'axe principal et au pourcentage d'explication des variables pour chaque axe résultant de la PCA. Si ceux-ci sont suffisamment grands comparés aux autres, cela veut dire que le groupe de variables explique bien le même phénomène d'où l'existence de cette corrélation. Ce procédé permet de posséder au final, un jeu de variables non corrélées entre elles et étant les plus explicatives des différents phénomènes ressortant de celles-ci. Concernant l'algorithme MaxEnt, quatre modèles différents sont construits pour chaque espèce avec des variables différentes. Celles-ci sont sélectionnées par rapport à leur contribution au sein du premier modèle qui regroupe toutes les variables dernièrement sélectionnées. Ainsi, les trois autres modèles sont construits sur base des variables dépassant un certain seuil de participation. Ces seuils sont respectivement, 2%, 3% et 5%. Cette dernière étape est réalisée afin d'assurer l'obtention du meilleur modèle possible mais la première reste la plus importante. C'est pourquoi, lors de l'utilisation de GAM, l'entièreté des variables non corrélées entre elles est utilisée exclusivement.

### 2.4.2. Création du background (Pseudo-absences)

Suite à l'absence de consensus scientifique sur la manière de définir les pseudo-absences, quatre méthodes différentes doivent être testées au préalable sur dix espèces afin de déterminer le procédé le mieux adapté à la situation. Celles-ci sont décrites successivement au sein du tableau 3. Grâce à elle, il est possible d'analyser l'effet de l'amélioration de la sélection des points et l'effet du nombre de points d'absences utilisés dans l'algorithme MaxEnt.

Tableau 3: Liste des méthodes testées pour la sélection des pseudo-absences (background) lors de l'utilisation de l'algorithme de MaxEnt.

Méthodes	Prévalence	Explication pratique
1. Aléatoire	$\neq 0,5$	10 000 points aléatoires différents des présences au sein de l'entièreté de la zone d'étude
2. Basé sur le nombre d'observations	$\neq 0,5$	Sites ayant plus de 6 observations et ne possédant pas l'espèce modélisée
3. Basé sur la richesse spécifique	$\neq 0,5$	Sites ayant plus de 3 espèces différentes recensées et ne possédant pas l'espèce modélisée
4. Basé sur la richesse spécifique	0,5	

L'utilisation du nombre d'observations ou la richesse spécifique au sein d'un carré kilométrique permettent d'estimer l'effort d'échantillonnage qui y a eu lieu et ainsi favoriser la sélection des absences dans des sites ayant eu un plus haut effort. Lors de ces comptages, les trois espèces aux points de présence particuliers n'ont pas été prises en compte car il ne s'agit pas véritablement de points d'observation brute. Par ailleurs, les deux seuils (de 3 et 6 espèces) ont été déterminés par essai erreur afin d'obtenir un nombre de carrés relativement grand et qui recouvre l'entièreté de la région Wallonne. Ce dernier point est assez important car chaque zone biogéographique doit être correctement représentée.

Par la suite, l'algorithme GAM est également testé en utilisant les pseudo-absences améliorées. L'objectif est de mettre en avant la qualité des absences ainsi sélectionnées en les utilisant avec un algorithme fonctionnant idéalement avec de vraies absences pour finalement comparer les résultats avec les méthodes précédentes.

#### 2.4.3. Sélection de la méthode

Dix espèces sont sélectionnées afin d'englober le plus de variabilité possible et donc de prendre en compte la majorité des différents cas de figure possibles. Par exemple, l'Autour des palombes est choisi car il s'agit d'une espèce ayant une aire de distribution assez grande et est par conséquent plus difficile à modéliser à l'opposé de la Grèbe huppé qui est une espèce plus spécialisée. Également, des espèces ayant de nombreux points d'occurrence et inversement sont sélectionnées.

La modélisation de chacune d'entre elles est réalisée une fois pour chaque différente méthode de sélection du background avec MaxEnt. GAM est testée pour chaque espèce une fois avec les deux différentes méthodes d'amélioration du background avec une prévalence inférieure à 0,5 (Tableau 3).

#### 2.4.4. Modélisation par MaxEnt

La modélisation nécessite au préalable une séparation du jeu de données en deux groupes de présences « indépendantes » qui sont les données de validation et calibration. Ces dernières servant à construire le modèle représentent 75% des données brutes alors que celles de validation en représentent 25% (Johnston et al., 2020). La même distinction est faite pour les points d'absence. Comme la résolution utilisée est supérieure à la précision de nombreux points de présence, plusieurs observations d'une même espèce peuvent se retrouver au sein du même carré. Or, les points de présence se résument tous au point central du carré kilométrique dans lequel il se trouve. De cette manière, un seul point de présence est considéré par carré kilométrique.

Cependant, même si l'algorithme de MaxEnt permet une modélisation avec peu de données de présence (Pearson et al., 2007), elle nécessite tout de même au moins 30 observations (Stockwell et al., 2002). Guisan et al. (2017) conseillent quant à eux, l'utilisation d'un échantillon d'au moins 50 observations. Ainsi, pour éviter de présenter des modèles fortement biaisés, les espèces pour lesquelles ce nombre au sein des données de calibration n'est pas atteint ne sont pas modélisées. GAM en est encore plus sensible et donc permettrait la modélisation de moins d'espèces s'il est finalement choisi. La liste des espèces dont leur distribution ne peut être modélisée se trouve en annexe (Annexe 3).

#### 2.4.5. Évaluation et validation des modèles

Comme il est expliqué lors de la sélection des variables, quatre modèles sont construits par espèce et le meilleur d'entre eux est sélectionné. Pour cela, cinq indices différents sont utilisés pour chaque modèle construit via MaxEnt qui sont l'AICc (Warren et al., 2011), l'AUC (A. Hanley et al., 1982; Osborne et al., 2002; Rosset, 2004), le Producer accuracy (PA) (Story et al., 1986), le User accuracy (UA) (Fung et al., 1988) et le coefficient de corrélation de Matthew (Chicco et al., 2020). Le premier permet de mesurer la complexité d'un modèle et son utilisation dans R a été développée par Muscarella et al., (2014). Le second évalue la capacité discriminative du modèle entre les absences et les présences (Fourcade et al., 2014). Celui-ci est mesuré en utilisant les données de validation sélectionnées précédemment. Ensuite, via la courbe ROC, le seuil de différenciation entre une présence et une absence est déterminé. Il s'agit du seuil qui maximise la sensibilité et la spécificité du modèle (Liu et al., 2013). Via celui-ci, les résultats proposés par MaxEnt sont transformés en données binaires afin de mesurer les autres indices. La sélection est réalisée en regardant leur valeur à tous, tout en mettant une importance supérieure à l'indice de Matthew et au UA afin de minimiser l'erreur de commission (fausses présences). Ces indices et leur formule sont repris dans le tableau 4.

Tableau 4: Indices d'évaluation des modèles sur base d'une matrice de confusion ainsi que leurs formules.

Indices	Formules
Coefficient de corrélation de Matthews	$\frac{VP \times VN - FP \times FN}{\sqrt{(VP + FP) \times (VP + FN) \times (VN + FP) \times (VN + FN)}}$
User Accuracy	$\frac{VP}{VP + FP}$
Producer Accuracy	$\frac{VP}{\text{Points de présence}}$

Pour finir, le modèle final est soumis à une validation croisée (Merow et al., 2013). Celle-ci consiste à élaborer à partir du jeu de données total, cinq sous-groupes aléatoires de présences et de pseudo-absences. Un modèle est construit sur base du même set de variables et en utilisant comme données de calibration toutes celles disponibles moins celles d'un groupe. Ces dernières sont utilisées en données de validation pour mesurer l'AUC. En résultat, cinq AUC différents sont déterminés et l'écart-type entre eux est mesuré. Si celui-ci est suffisamment faible, le modèle est validé. En effet, l'AUC possède également comme défaut de ne pas parvenir à informer sur l'existence d'un éventuel problème de discrimination locale (Lobo et al., 2008). Via cette validation croisée, il est donc possible de déterminer plus ou moins l'existence de différence géographique importante au sein du jeu de données. Pour finir, cela permet de donner également un écart-type aux autres indices lors de la sélection de la méthode pour comparer leur valeur entre deux méthodes différentes.

#### 2.4.6. Comparaison et choix de la méthode finale

Les différents indices mesurés pour l'évaluation de chaque modèle ainsi que la carte résultante sont comparés par espèce afin de sélectionner la meilleure méthode. Ensuite, celle finalement choisie est utilisée de la même manière que pour les dix espèces tests mais sur l'entièreté des espèces rares. Il s'agit de la méthode par défaut de MaxEnt.

### 2.5. *Sélection des espèces indicatrices*

En Wallonie, l'un des gradients les plus importants pouvant influencer la présence ou l'absence d'une espèce est bien l'altitude (Dufrene et al., 1991; Brogna et al., 2018). Or, l'objectif ici est de déterminer les espèces liées spécifiquement à une ou plusieurs trames représentées par des occupations du sol. De ce fait, l'analyse des modèles est réalisée pour chaque région biogéographique afin de mettre le plus en avant l'effet de l'occupation du sol sur la présence ou l'absence des espèces. Il s'agit des mêmes territoires biogéographiques présentés précédemment à l'exception où le Condroz, le territoire hennuyer-brabançon-Hesbignon et le sillon Sambre-mosan sont regroupés. Ce choix pourrait sembler anormal à première vue mais cela permet de diminuer le nombre d'analyses et le gradient d'altitude entre ces deux régions n'est pas aussi fort qu'avec l'Ardenne. La Fagne-Famenne est également considérée à part entière car elle forme une dépression entre l'Ardenne et le Condroz, ce qui a plus de chance d'offrir certaines particularités dans la distribution des espèces. Grâce à cela, la variance expliquée et l'influence de l'occupation du sol sont mieux visibles et plus fortes.

Une fois de plus, le travail s'est réalisé séparément entre les espèces communes et rares tout en maintenant un procédé similaire. Chaque trame écologique est caractérisée par une ou plusieurs variables d'occupations du sol (Tableau 2). Ainsi, par région, la valeur des pixels de ces variables rastérisées précédemment est extraite ainsi que celle des rasters résultants de MaxEnt pour les espèces rares et des modèles d'abondances pour les autres. Cela donne dans les deux cas, deux matrices quantitatives représentant d'un côté la proportion de chaque occupation du sol au sein d'un carré kilométrique et de l'autre, le nombre d'individus ou la probabilité de présence de l'espèce. Pour l'entièreté de celles-ci, la distribution de leur probabilité d'occurrence ou de leur abondance n'est pas normale mais en « j inversé ». Aucune transformation ne semble cependant arranger le problème. Cependant, ces distributions restent décroissantes, monotones et continues sans outliers. Pour améliorer l'analyse, les espèces rares dont les gammes de valeurs s'éloignent de trop de [0 ;1] sont mises à l'écart, ce qui permet d'éviter l'obtention de résultats aberrants. L'analyse mise en œuvre est basée sur des RDA et se réalise de la même façon pour les deux groupes. Par région biogéographique et par groupe, une RDA est réalisée en prenant en compte l'ensemble des variables afin d'avoir une idée d'ensemble et vérifier si les trames et les espèces se placent correctement au sein du graph. Ceci est répété par la suite mais en prenant uniquement en compte les variables liées à une trame, principale ou secondaire et en construisant un cercle de corrélation représentant toutes les occupations du sol (trames). Pour finir, les espèces sont sélectionnées par trame en fonction de leur  $\cos^2$  (pourcentage de leur variance expliquée par l'axe), leur écologie et leurs interactions avec les autres trames en analysant les graphiques et surtout les cercles de corrélation. Seules les espèces communes ayant un  $\cos^2$  supérieur à 25% sont maintenues comme espèces potentiellement indicatrices avant de vérifier les autres critères. Ce seuil diminue à 10% pour les espèces rares. Le nombre cible d'espèces indicatrices par trame se situe entre 2 ou 3 pour les espèces rares et entre 3 et 4 pour les communes.

Concernant spécifiquement la variable « bocages », celle-ci ne provient pas de la couche écotope. Elle a été créée antérieurement à ce travail par photo-interprétation par Axel Bourdouxhe<sup>2</sup>, donnant une couche vectorielle représentant les limites d'un bocage. La surface de ces polygones au sein de chaque carré kilométrique est mesurée à l'aide d'Arcmap afin d'obtenir le même genre de cartes que celles produites à partir des écotopes. Elle est ensuite, multipliée par mille pour que sa proportion au sein de chaque carré soit exprimée de la même façon que les autres occupations du sol. C'est pourquoi, la trame « pseudo-lisière » composée des plantations, des forêts claires, des coupes forestières et trouées et la trame des bocages sont traitées séparément durant les analyses qui suivent.

## 2.6. Réseau écologique existant

Dans un premier temps, il est important de déterminer les zones cœurs déjà existantes pour être en accord avec la méthodologie de Dufrêne et al., (2019). Elle définit ces zones comme répondant aux critères suivants : une biodiversité observée importante, couverte par les unités de gestions prioritaires de Natura 2000, par la LCN ou encore par des SGIB. Ce réseau regroupe par conséquent l'ensemble des zones ayant une biodiversité remarquable reconnue (avifaune ou non) et ceux étant déjà sous protection. Afin de « matérialiser » ces critères, l'indice alpha (Chmura et al., 2006) sur la diversité non modélisée est calculé au sein de chaque carré kilométrique. Étant donné la non-accessibilité des données de présences brutes concernant les espèces communes, il ne reflète que la diversité spécifique des espèces rares. Ensuite, les shapefiles représentant tous les différents sites relatifs à la loi sur la conservation de la nature (réserves naturelles, réserve domaniale, ZHIB et réserves naturelles agréées) sans les parcs naturels ainsi que les unités de gestion du réseau Natura 2000 et les SGIB sont analysés. Grâce à ces couches vectorielles, une carte à résolution d'1km<sup>2</sup> peut être construite afin de représenter les proportions surfaciques de chaque site. Ceci génère trois couches rasters représentant leur densité surfacique dans un carré kilométrique. Elles sont toutes les trois additionnées avec une pondération triple pour les sites sous la LCN et double pour les SGIB. Les valeurs des pixels de ce raster sont séparées en deux groupes les plus homogènes via la fonction de classification proposée par ArcMap 10.7.1 de Jenks (Natural breaks). Cette dernière permet de classer des valeurs en minimisant la variance au sein de ces classes et ainsi maximiser la différence entre elles. Ainsi, les pixels sont séparés en deux classes distinctes et la borne inférieure de la dernière classe est choisie comme seuil. L'indice alpha quant à lui doit être analysé séparément selon la région biogéographique de la même manière que lors de la détermination des espèces indicatrices. Le seul changement est que le Condroz et le nord du sillon Sambre et Meuse restent séparés. Pour se faire, l'indice alpha est classé par région en quatre classes et le seuil vaut la valeur de la borne supérieure de la deuxième. Ces deux cartes binaires sont ensuite sommées pour donner une représentation cartographique du réseau écologique existant. Cependant, il est encore plus intéressant de créer une carte spécifique à la LCN afin de posséder d'un côté les sites déjà sous protection et de l'autre, les sites ayant uniquement une biodiversité élevée reconnue. Pour cela, un seuil spécifique aux sites sous protection de la LCN est déterminé une fois encore par classification binaire de Jenks. Il en est de même pour les SGIB dont la carte résultante est additionnée à celle de l'indice alpha observé.

---

<sup>2</sup> Doctorant à la faculté de Gembloux Agro-Bio Tech, ULiège, au sein du département Biodiversité et Paysage travaillant pour le projet LifeWatch



## 2.7. Réseau écologique potentiel

Grâce aux rasters de présence des espèces indicatrices sélectionnées au point 2.5, il est possible d'identifier l'appartenance d'un carré kilométrique à une trame. Cependant, se limiter à la simple présence d'une seule espèce indicatrice pour la désignation d'une zone cœur de biodiversité n'est pas suffisant. Pour cela, il faut combiner ces présences avec l'indice alpha modélisé. Celui-ci est la mesure des présences estimées de toutes les espèces qu'elles soient classées comme rares ou communes. Ainsi, il permettra d'identifier les carrés kilométriques ayant une diversité estimée importante indépendamment de la trame. Pour se faire, un seuil doit être déterminé afin de différencier un carré kilométrique ayant une bonne biodiversité des autres de la même manière que pour l'indice alpha observé. Une fois de plus, une analyse par région biogéographique est réalisée afin d'obtenir les seuils les plus pertinents.

Par conséquent, une zone cœur doit respecter ces deux critères de sélection imposés par le seuil de présences modélisées et de la présence d'au moins une espèce indicatrice. De cette manière, seuls les carrés kilométriques ayant une bonne biodiversité en oiseaux et représentant la trame correspondante sont sélectionnés. Cependant, il semblerait que cette première méthode testée ne fonctionne pas assez bien et provoque la sélection d'un trop grand nombre de zones cœurs.

Pour pallier à ce problème, une nouvelle couche raster est construite par trame écologique. Elles représentent concrètement la diversité alpha en espèces rares pondérée par les résultats des RDA réalisées au point précédent. Ainsi, la présence d'une espèce prenant la valeur de 1 est multipliée par sa coordonnée relative à l'axe de la trame en question. Cela est fait ainsi avec chaque espèce rare pour toutes les trames. Les rasters résultants sont par la suite transformés en rasters binaires dont le seuil de différenciation est également déterminé via la classification de Jenks. Cette classification se fait par région biogéographique pour chaque trame. Plus précisément, les valeurs des pixels d'une trame dans une région sont regroupées en trois classes et donc avec six bornes via la fonction R « `getJenksBreaks` » du package « `BAMMtools` ». Le seuil vaut la valeur de la cinquième borne. Celles qui sont négatives sont amenées à zéro.

Grâce à cette dernière sélection, les carrés kilométriques représentant proportionnellement mieux une trame par région sont favorisés. Ceci devrait augmenter la proportion de l'occupation du sol de la trame dans son réseau.

Les corridors reliant ces différentes zones cœurs sont délimités via le logiciel Graphab (Foltête et al., 2012). Celui-ci permet d'identifier les chemins les plus courts possible entre deux taches en se basant sur la théorie des graphes. Il permet aussi de déterminer les chemins ayant les plus faibles coûts pour l'espèce mais la résolution de travail étant trop grossière, seules les distances euclidiennes sont utilisées. Trois distances de dispersion moyennes sont choisies afin d'englober la majorité des distances possibles au sein de l'avifaune. Il s'agit des distances de 2, 5 et 10km. Concrètement dans le logiciel, la taille minimale des habitats est fixée à 100ha et les taches sont simplifiées afin d'offrir un graphe planaire limitant le nombre de liens d'une tache à quatre maximum. Ensuite, un graphique seuillé est généré avec les trois différentes distances. Finalement, deux indices globaux sont calculés par graphe. Il s'agit de la Probabilité de Connexion et de l'indice d'Harary dont les formules sont reprises dans le tableau 5. L'indice de comparaison nommé dPC est également mesuré au sein de chaque lien des trois graphes pour chaque zone centrale et pour chaque distance. Les distances intrataches ne sont pas prises en compte dans la mesure des métriques car le coût éventuel de déplacement d'une espèce au sein d'une zone cœur de sa trame est considéré comme nul.

Tableau 5: Indice d'évaluation d'un réseau associé à leur formule et leurs bornes supérieures et inférieures. \*La valeur de d est égale à la distance de dispersion moyenne choisie (2,5 ou 10km).

Indice	Formule	Borne
Probabilité de connexion	$PC = \frac{1}{A^2} \sum_{i=1}^n \sum_{j=1}^n a_i * a_j * e^{-\alpha d_{ij}}$ <p> <math>a = -\frac{\log(p)}{d}</math> où <math>p = 0.5</math> et <math>d^*</math>  <math>\alpha =</math>  <math>d_{ij}</math> = distance inter-tache  <math>A</math> = Surface totale de toutes les taches </p>	[0 ;1]
Indice d'Harary	$H = \frac{1}{2} * \sum_{i=1}^n \sum_{j=1}^n \frac{1}{n * l_{ij}}$ <p>Où <math>j \neq i</math> et <math>n</math> = nombre de taches</p>	$[0 ; \frac{n*(n-1)}{2}]$

Les liens et les taches générés au sein de Graphab sont exportés sur SIG pour des manipulations supplémentaires. Tout d'abord, tous les liens ayant un dPC nul sont supprimés car ils n'apportent rien à la connectivité du réseau. Ensuite, deux cartes de chaleurs sont générées par jeu de liens via le logiciel QGIS 2.18.24. La première est construite via les points centraux de chaque lien. La deuxième suit le même principe sauf qu'elle est pondérée par le dPC connector du lien correspondant. De cette manière, le premier raster indique le réseau écologique en lui-même sur deux dimensions et le second fait de même en indiquant les zones les plus importantes pour la connectivité de ce réseau écologique.

Une dernière représentation cartographique est réalisée afin de mettre en évidence pour chaque réseau, les zones cœurs ayant une déjà une biodiversité reconnue ainsi que ceux d'ores et déjà sous protection. Ces dernières sont déjà représentées via le réseau écologique existant. Pour obtenir cela, les deux cartes générées à la fin du point 2.6 sont croisées avec le réseau écologique final de chaque trame pour mettre en évidence les carrés kilométriques communs.

## 2.8. Analyse des réseaux écologiques

Les réseaux sont tout d'abord additionnés afin de créer un réseau global et de mesurer le nombre de fois qu'un carré kilométrique est repris comme zone cœur. Ensuite, une comparaison deux à deux de chaque réseau est réalisée via la création d'un tableau à deux entrées. Dans celui-ci sont mesurées les proportions de surfaces partagées entre deux réseaux de trames différentes uniquement via les zones cœurs.

Pour ce qui est de l'évaluation plus concrète des réseaux, la surface totale de chaque occupation du sol représentant une trame écologique est mesurée et comparée avec la surface totale du réseau correspondant. Afin d'apporter plus de pertinence à ces mesures, celles-ci ont une fois de plus été réalisées séparément par région biogéographique. De cette façon, la proportion de la trame effectivement recouverte par son réseau est mesurée. Également, une approximation de la « précision » de chaque réseau est réalisée via la mesure de la proportion surfacique qu'occupe une trame au sein de son propre réseau. Par la suite, la mesure de la moyenne surfacique de l'occupation du sol de chaque trame est mesurée par région ainsi que celle lors qu'uniquement les carrés kilométriques désignés comme ZC sont pris en compte. De cette manière, il est possible de calculer d'une autre façon la précision des réseaux écologiques.

Pour compléter l'évaluation concrète de chaque réseau, une analyse plus approfondie est faite par rapport à la diversité modélisée et observée au sein des zones cœurs. Les moyennes par carré kilométrique de chaque réseau et pour les deux indices sont calculées et comparées avec celle obtenue sur l'ensemble de la Wallonie. Également, le nombre d'observations par carré kilométrique est réutilisé afin d'estimer la popularité ornithologique d'un site et de comparer celle en dehors et à l'intérieur des zones cœurs du réseau global. Pour cela, un simple test khi carré est réalisé afin de déterminer une dépendance entre la popularité d'un site et sa présence dans un réseau. Pour cela, les sites doivent être classés en deux catégories, populaires et non populaires. Encore une fois, la classification de Jenks a permis cette séparation en utilisant cette fois-ci 6 classes et en prenant comme seuil, la borne supérieure de la deuxième classe.

L'hypothèse initiale est que la mise sous réseau via cette méthodologie, permet une augmentation de la richesse spécifique moyenne des carrés. Dans un second temps, l'analyse du nombre d'observations détermine si les carrés populaires sont effectivement repris au sein du réseau global. Dans cette optique, il est important d'identifier les autres variables influençant la popularité et la richesse spécifique d'un milieu. Deux variables doivent être contrôlées qui sont l'altitude et l'occupation du sol. Pour la première, les calculs de la moyenne sont mesurés par biogéorégion de la même manière que précédemment. Pour la seconde variable, trois matrices de corrélation sont construites entre la distribution des trois indices et les différentes occupations du sol afin d'identifier leurs éventuelles interactions. Ces analyses sont également réalisées séparément par régions.

Le dernier point le plus important à réaliser est bien entendu de comparer les réseaux potentiels et le réseau existant. À nouveau, il est plus intéressant de comparer les zones d'ores et déjà protégées, celles ayant une biodiversité remarquable connue et les zones cœurs potentielles. Cette analyse a déjà été initiée en fin du point 2.7 mais cette fois-ci, il s'agit directement des surfaces qui sont comparées éliminant l'éventuel biais apporté par un seuil. Ces surfaces correspondent aux SGIB et aux sites sous la LCN. De cette manière, deux fois deux proportions sont calculées. La première correspond à la proportion de ces sites se situant au sein du réseau écologique global par rapport à la superficie totale sous protection. Le second vaut la proportion de sites nommés et en même temps couverts par un réseau par rapport à la surface totale des réseaux. Concernant la biodiversité observée, les seuils déterminés précédemment différenciant les carrés kilométriques ayant une bonne biodiversité observée des autres sont réutilisés. Grâce à la cartographie binaire de l'indice alpha observé, la proportion de carrés kilométriques au sein du réseau global et ayant une diversité observée remarquable peut être mesurée. Cette analyse est une variante de celle réalisée au paragraphe précédent concernant ce même indice.

Finalement, un bref rapport peut être réalisé sur la richesse de l'avifaune en Wallonie et dans ses différentes régions via les deux indices alpha mesurés et ceci, indépendamment des réseaux écologiques.

Afin de bien illustrer la globalité de la méthode de construction d'un réseau, la figure 1 regroupe l'entièreté des étapes présentées en partant des données d'observations brutes jusqu'à l'obtention des réseaux écologiques potentiels finaux. La figure 2 illustre les analyses des réseaux jusqu'à l'analyse des indices de diversité. Les cartes qui y sont affichées sont présentées bien plus en détail dans la partie « Résultats ». Également, pour l'ensemble des cartes affichant le réseau d'une trame spécifique, la couche écotope la représentant sera systématiquement affichée avec elle. La graduation de ces dernières correspond à la surface en km<sup>2</sup> de l'occupation du sol multipliée par 1000.

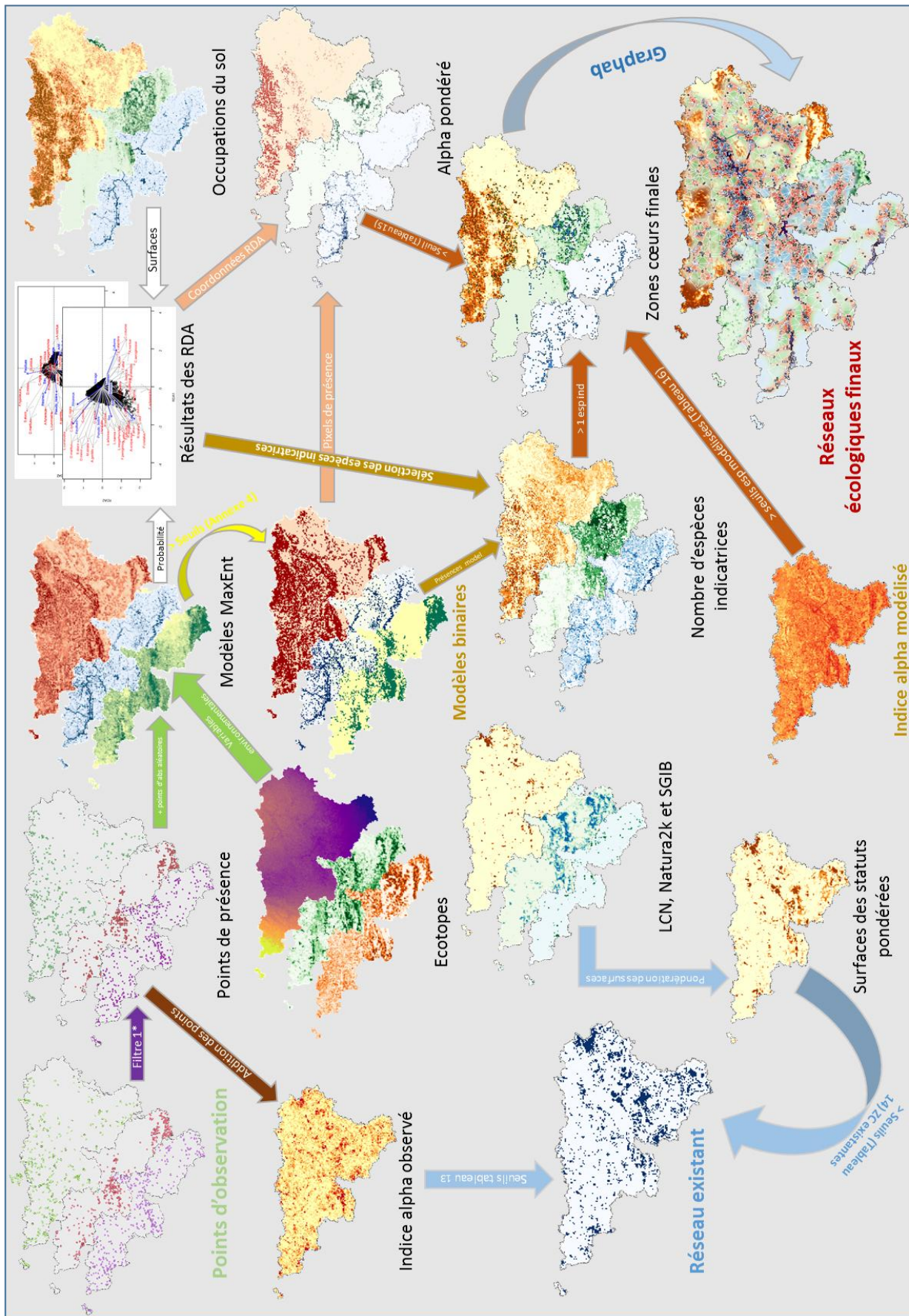


Figure 1: Schéma reprenant les points essentiels de la méthodologie de sélection des espèces indicatrices, de désignation des zones cœurs de biodiversité ainsi que de la délimitation des réseaux écologiques. Les flèches directives de même couleur décrivent de façon complémentaire le processus de création de la carte désigné par celles-ci. Seul le lien entre la carte représentant l'indice alpha modélisé et les rasters de présences modélisées n'a pas pu être représenté. \*Filtre 1 réfère au point 2.2 "Gestion des données" où les points n'ayant pas une précision suffisante et ceux se trouvant en dehors de la Région Wallonne ont été supprimés.



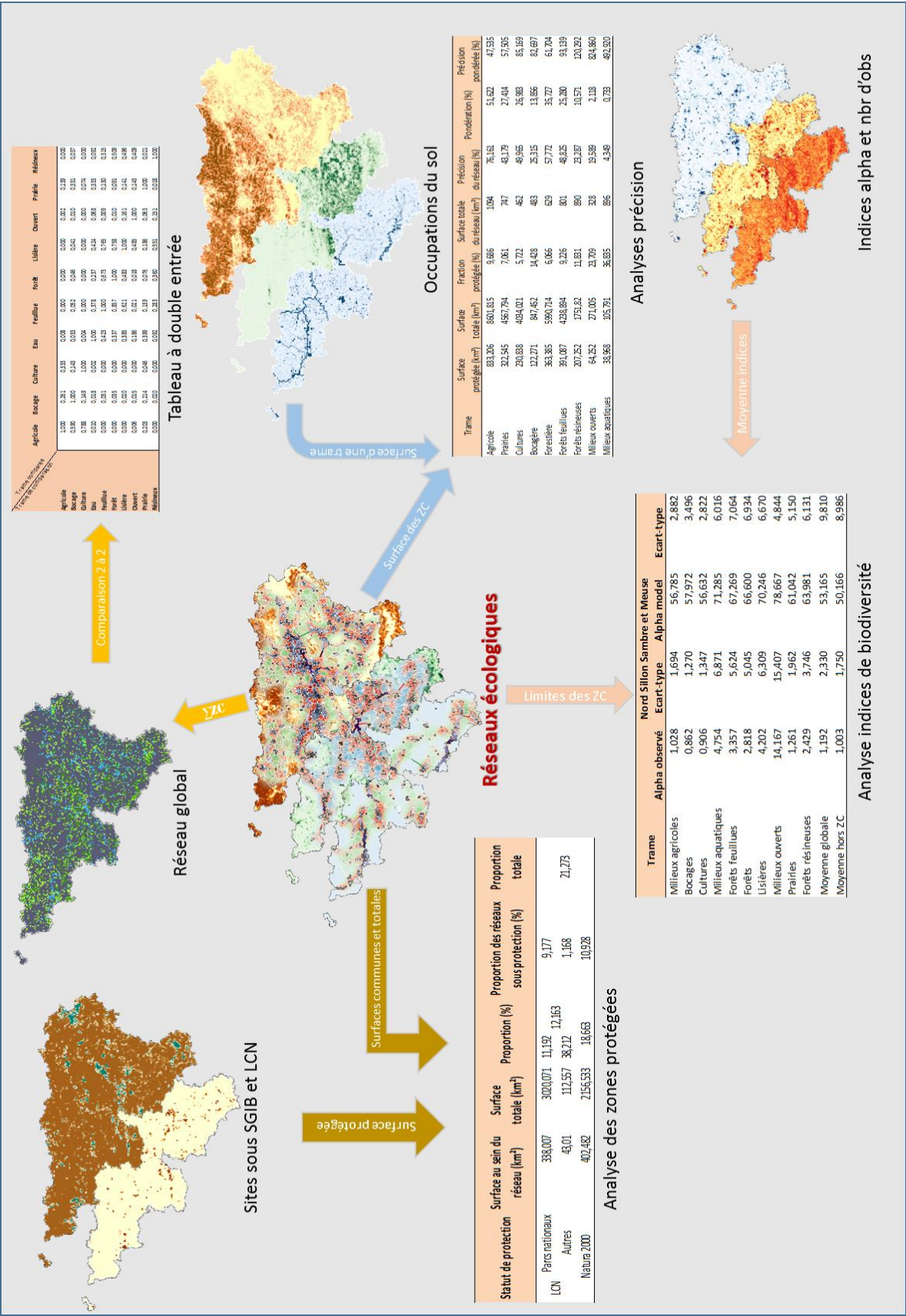


Figure 2: Schéma reprenant les analyses principales réalisées après la construction des différents réseaux : analyse de la précision, des indices de diversité au sein des réseaux et leur croisement avec le réseau existant ainsi que leur comparaison deux à deux.

### III. Résultats

#### 3.1. Révisions des trames écologiques

Après analyse des trames proposées par Dufrêne et al., 2019 et révisions de la littérature, les trames initiales ont été modifiées pour mieux s'adapter à l'avifaune. Aucune des trames principales initiales n'a été cependant supprimée. Le tableau 6 présente la liste des trames principales et secondaires utilisées lors du classement manuel de chaque espèce au sein d'une trame ainsi que leur fréquence d'apparition lors de ce classement. La fréquence totale est supérieure au nombre total d'espèces étudiées car la grande majorité est classable dans plusieurs trames écologiques. La quasi-totalité des trames secondaires proposées à la suite de cette analyse n'a pas été gardée en raison de leur fréquence inférieure à 1%. Seule la trame des forêts feuillues a subi des changements notables tels que la suppression des forêts sur sol sec et mésophile ainsi que la création de la sous-trame « forêt claire ». Également, trois trames supplémentaires ont été ajoutées. Il y a tout d'abord les forêts résineuses qui semblent avoir une importance assez considérable pour l'avifaune et plus particulièrement pour une trentaine d'espèces. Parmi elles, peuvent être cités le Gros-bec casse-noyaux et le Bec-croisé des sapins. Cette nouvelle trame se divise de la même façon que celle des forêts feuillues, à savoir, les vieilles forêts (sombre, dense et avec des strates inférieures peu développées) et forêts claires (avec stratification développée, jeunes et peu denses). Ensuite, de nombreuses espèces se sont fortement liées au cours de leur évolution au caractère anthropique de certains milieux, c'est pourquoi, une trame des milieux artificiels a été créée. Cette dernière regroupe les milieux urbains, les bâtiments et infrastructures isolées, les jardins et les parcs enclavés en ville. Une trentaine d'espèces sont liées à celle-ci comme le Martinet noir, l'Hirondelle rustique et la Bergeronnette de Yarrell. La dernière trame ajoutée est celle des milieux rocheux qui concerne une petite dizaine d'espèces telles que l'Hirondelle des rivages et fenêtres et le Faucon pèlerin. Les autres trames n'ont pas été modifiées sauf celle des bocages et lisières qui a été affinée en trois sous-trames. Elle représente donc les réseaux de haies, les plantations et les bosquets. Elles sont toutes présentées au sein de la figure 3 avec leurs sous-trames. Concernant celles des milieux forestiers, une trame supérieure est ajoutée afin de représenter les forêts mixtes. Celle-ci n'est pas présente dans le tableau 6. Suivant la même idée, une sous-trame opposée aux cultures a été ajoutée qui est celle des prairies.

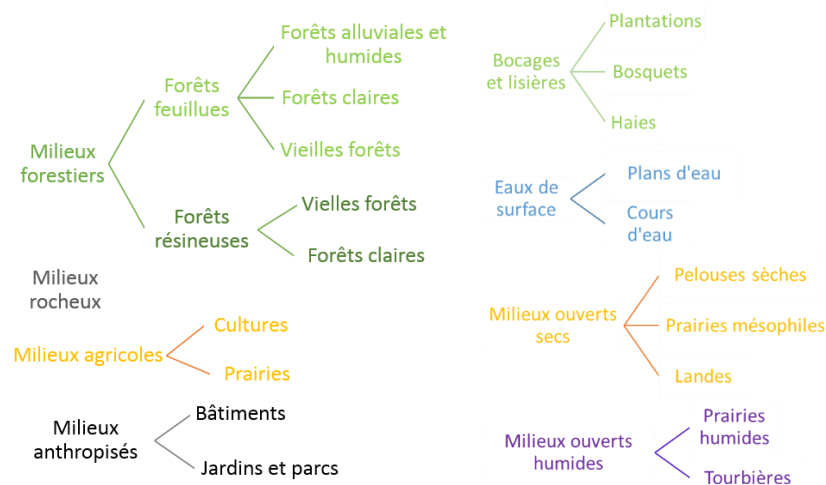


Figure 3: Trames écologiques redéfinies afin de mieux englober les demandes écologiques, abiotiques et biotiques de l'avifaune en Région Wallonne.

## Résultats

Tableau 6: Liste des trames candidates à la composition du réseau écologique sur l'avifaune associées à leur fréquence relative, les trames secondaires retenues malgré leur fréquence inférieure à 1% sont indiquées en vert et celles non maintenues sont en rouge, 1% vaut 4 espèces.

Trames proposées	Fréquences relatives	Maintient
<b>Milieux ouverts secs</b>	58	OUI
Pelouses sèches	0	NON
Prairies mésophiles	2	NON
Landes	1	NON
<b>Milieux semi-ouverts</b>	1	NON
<b>Eaux de surface</b>	26	OUI
Cours d'eau	4	OUI
Plans d'eau	9	OUI
<b>Forêts feuillues</b>	47	OUI
Forêts sur sol sec	0	NON
Forêts mésophiles	1	NON
Forêts alluviales et humides	18	OUI
Forêts claires	14	OUI
Forêts jeunes	2	NON
Forêts résineuses	30	OUI
Forêts mixtes	2	NON
<b>Bocages et lisières</b>	62	OUI
<b>Milieux agricoles</b>	20	OUI
Cultures	6	OUI
<b>Milieux anthropisés</b>	35	OUI
<b>Milieux rocheux</b>	6	OUI
<b>Milieux ouverts humides</b>	39	OUI
Prairies humides	8	OUI
<b>Tourbières</b>	2	NON

### 3.2. Modélisation

Les meilleures variables écotopes ayant été maintenues pour la modélisation sont présentées au sein du tableau 7. Certaines variables ont été enlevées manuellement telles que « Light » naturellement corrélée à « Artif ». La première représente le taux de lumières artificielles et la seconde, l'occupation anthropique du sol mais leur corrélation n'est pas détectée lors de l'analyse. L'azimut ainsi que la pente en degré sont également mis de côté car ces variables sont corrélées à d'autres de meilleure qualité ce qui justifie leur exclusion.

Tableau 7: Liste des variables écotopes utilisées pour la modélisation classée selon leur différent type.

Wordclim	Lifewatch-WB	Topographie	Occupation du sol	Attribut du sol	Profondeur du sol	Drainage
Antot_P	SnowEnd	Slope_prc	Artif	Allupegble	Dpt_deep	Dr_dry
STD_T	SnowLength	SunSpring	Bare	Allusilt	Dpt_superf	Dr_mid
MDryQ_T			BroadL	Clay	Dpt_zero	Very_humid
MinColdQ_T			Copen	Organic		
MWetQ_T			Mgramin	Podzol		
			NeedL	Calcareous		
			Nopen			
			Plowed			
			Source			
			Water			
			Wopen			

Ensuite, dix espèces ont été sélectionnées afin de tester les quatre méthodes proposées précédemment. Le résultat des différents indices ainsi que leur écart-type de chaque espèce sont présentés au sein du tableau 8. Il s'agit pour rappel de l'AICc, l'AUC, du Producer Accuracy (PA), du User Accuracy (UA) et du coefficient de corrélation de Matthew. En raison des résultats très mauvais obtenus lors de l'égalisation des prévalences, les résultats de celles-ci ne sont pas présentés. Les modèles comparés n'ont pas forcément le même seuil de participation de leurs variables environnementales. À chaque fois, le meilleur modèle pouvant être tiré de chaque méthode est utilisé pour la comparaison afin qu'elle soit la plus pertinente possible. Il peut donc y avoir un ou plusieurs jeux de variables différents pour une même espèce. Les cartes résultantes du **Pic épeichette** et du **Petit gravelot** se trouvent en figure 4.

## Résultats

Tableau 8: Résultats des trois méthodes de sélection du background concernant la modélisation via l'algorithme MaxEnt. Elles réfèrent respectivement à la sélection de 10 000 points aléatoires, de l'utilisation du nombre d'espèces recensé et l'utilisation du nombre d'individus observés comme proxy de l'effort d'échantillonnage.

Espèce	Méthode	Modèle	Data_cal	Data_val	Abs_cal	Abs_val	AICc	AUC	Matt	PA	UA
<i>Accipiter gentilis</i>	1	2	384	127	7500	2500	7753,801±29,752	0,786±0,018	0,259±0,034	0,625±0,087	0,171±0,039
	2	1	384	127	829	276	8409,303±5,713	0,624 ±0,038	0,393±0,056	0,880±0,087	0,472±0,067
	3	1	384	127	1102	368	8321,216±50,826	0,663±0,003	0,374±0,049	0,916±0,037	0,392±0,037
<i>Asio otus</i>	1	2	471	156	7500	2500	9795,384±38,12	0,622±0,0315	0,120±0,009	0,73 ±0,118	0,088±0,009
	2	1	471	156	996	332	10186,63±23,403	0,732±0,015	0,487±0,029	0,8123±0,049	0,568±0,038
	3	2	471	156	1217	406	9918,495±13,280	0,692±0,030	0,332±0,017	0,749±0,121	0,437±0,050
<i>Dendrocopos minor</i>	1	1	487	161	7500	2500	9988,764±69,770	0,795±0,014	0,280±0,019	0,081±0,046	0,163±0,019
	2	1	487	161	934	318	10215,470±32,967	0,758±0,267	0,516±0,335	0,837±0,071	0,597±0,065
	3	3	487	161	1187	396	9807,531±16,339	0,763±0,017	0,370±0,026	0,764±0,112	0,469±0,041
<i>Falco subbuteo</i>	1	2	97	32	7500	2500	/	0,695±0,0235	0,167±0,053	0,816±0,119	0,055±0,034
	2	4	97	32	1150	383	2099,355±20,528	0,595±0,049	0,198±0,044	0,566±0,309	0,218±0,131
	3	3	97	32	1376	458	/	0,671±0,034	0,276±0,057	0,827±0,122	0,170±0,122
<i>Podiceps cristatus</i>	1	2	114	37	7500	2500	1754,711±23,451	0,972±0,015	0,381±0,033	0,927±0,0214	0,171±0,030
	2	3	114	37	1130	377	1787,208±18,602	0,938±0,018	0,588±0,039	0,871±0,043	0,467±0,066
	3	2	114	37	1356	452	1782,953±19,742	0,948±0,010	0,582±0,032	0,882±0,033	0,443±0,056
<i>Rallus aquaticus</i>	1	1	56	18	7500	2500	/	0,910±0,043	0,181±0,023	0,959±0,015	0,042±0,009
	2	3	56	18	1166	388	1210,297±87,127	0,862±0,036	0,361±0,057	0,845±0,051	0,205±0,057
	3	2	56	18	1382	461	/	0,883±0,038	0,376±0,048	0,871±0,137	0,211±0,088
<i>Cettia cetti</i>	1	4	68	22	7500	2500	1270,706±87,105	0,964±0,021	0,293±0,050	0,911±0,050	0,107±0,038
	2	2	68	22	1158	386	1303,365±23,350	0,961±0,011	0,502±0,055	0,928±0,082	0,324±0,092
	3	3	68	22	1379	460	1152,021±59,203	0,936±0,035	0,440±0,064	0,931±0,547	0,255±0,070
<i>Charadrius dubius</i>	1	1	44	14	7500	2500	/	0,913±0,031	0,142±0,021	0,996±0,009	0,026±0,007
	2	1	44	14	1180	393	/	0,865±0,052	0,376±0,124	0,944±0,049	0,198±0,091
	3	1	44	14	1397	466	/	0,875±0,401	0,290±0,054	0,991±0,019	0,115±0,030
<i>Corvus frugilegus</i>	1	2	321	107	7500	2500	6329,589±9,240	0,838±0,011	0,243±0,017	0,839±0,075	0,115±0,017
	2	2	321	107	1169	390	64006,824±29,305	0,905±0,011	0,597±0,023	0,861±0,041	0,570±0,044
	3	1	321	107	1346	448	6688,382±49,672	0,882±0,02	0,579±0,044	0,926±0,029	0,501±0,057
<i>Emberiza calandra</i>	1	3	210	69	7500	2500	3517,01±34,781	0,956±0,012	0,428±0,060	0,877±0,067	0,242±0,079
	2	2	210	69	1172	390	3700,396±11,533	0,978±0,004	0,808±0,049	0,913±0,097	0,775±0,097
	3	3	210	69	1328	443	3445,784±42,562	0,969±0,012	0,770±0,038	0,901±0,03	0,733±0,076



## Résultats

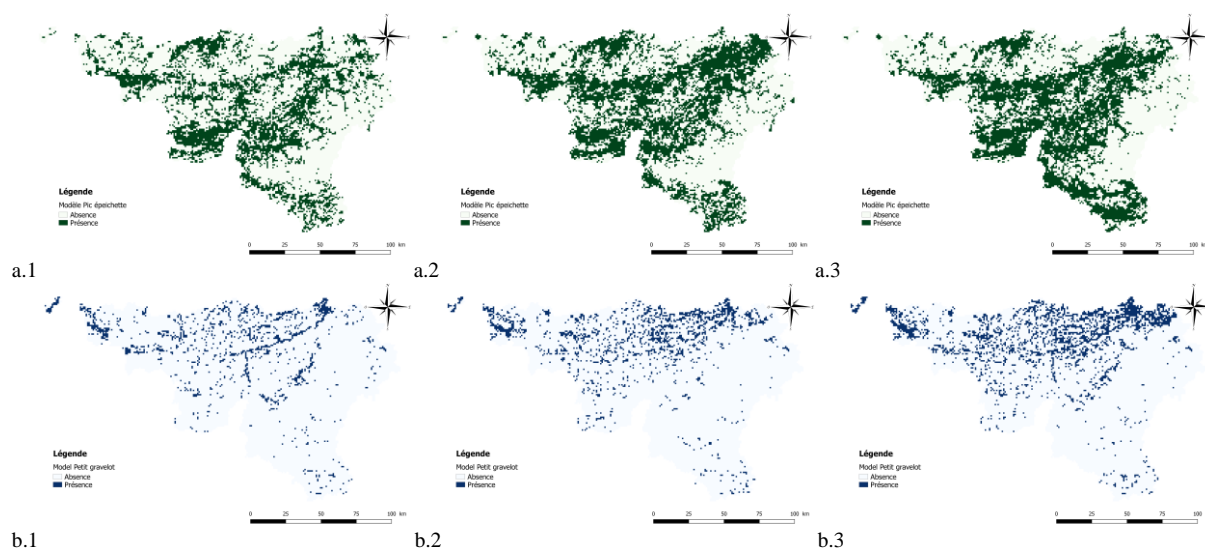


Figure 4: Cartes des présences/absences résultantes de la modélisation de deux espèces via les trois méthodes différentes. Le Pic épeichette (*Dendrocopos minor*) (a) et le Petit gravelot (*Charadrius dubius*) (b) y sont donc représentés. Les nombres 1, 2 et 3 réfèrent respectivement à la sélection de 10 000 points aléatoires, de l'utilisation du nombre d'espèces recensées et l'utilisation du nombre d'individus observés comme proxy de l'effort d'échantillonnage.

Du point de vue de l'AUC, la première méthode offre quatre fois sur dix des valeurs significativement supérieures aux autres et n'est que trois fois inférieure à au moins une autre méthode. Cependant, même si cette infériorité existe, elle n'apparaît que fortement pour le Hibou moyen duc. En ce qui concerne les autres indices par contre, elle présente des valeurs nettement inférieures aux autres sauf pour le PA où aucune réelle différence n'est observée. Il serait donc avisé à première vue, d'utiliser la méthode 2 qui possède globalement de meilleurs résultats tous indices confondus. Cependant, après observation des cartes de présences et absences (Figure 4), la méthode une et deux sont fortement généreuses dans leurs estimations. Afin d'éviter une surestimation des distributions spatiales des espèces rares, la première a finalement été sélectionnée pour la suite. Les modèles offerts par l'algorithme GAM dont les résultats des indices se trouvent dans le tableau 9 ne sont pas assez suffisants même s'ils donnent tout de même pour certaines espèces de bons résultats.

Tableau 9: Valeur des AUC des modèles construits via l'algorithme de GAM. La première méthode utilise la richesse spécifique d'un carré kilométrique comme proxy de l'effort d'échantillonnage et la seconde, le nombre d'observations. La modélisation via GAM a été réalisée par Thomas Coppée<sup>3</sup>.

Espèce	Méthode 1	Méthode2	MaxEnt
<i>Accipiter gentilis</i>	0,65	0,592	0,786
<i>Asio otus</i>	0,71	0,759	0,622
<i>Dendrocopos minor</i>	0,651	0,947	0,795
<i>Falco subbuteo</i>	0,752	0,783	0,695
<i>Podiceps cristatus</i>	0,887	0,907	0,972
<i>Rallus aquaticus</i>	0,749	0,781	0,91
<i>Cettia cetti</i>	0,893	0,931	0,964
<i>Charadrius dubius</i>	0,499	0,633	0,913
<i>Corvus frugilegus</i>	0,918	0,905	0,905
<i>Emberiza calandra</i>	0,724	0,691	0,956

<sup>3</sup> Doctorant à Gembloux Agro-Bio Tech au sein de l'unité Biodiversité et Paysage pour le projet LifeWatch.

Par la suite, tous les indices utilisés lors de cette sélection ont à nouveau été utilisés pour caractériser le modèle final de chaque espèce « rare » et leurs valeurs sont disponibles en annexe (Annexe 4). À chaque fois, la considération du UA et du coefficient de Matthew sont favorisées afin de réduire le plus possible les erreurs de commissions tout en ayant une précision globale assez bonne. Malgré tout, certains modèles ne sont relativement pas bons et doivent être utilisés avec précaution. Une série d'espèces a été classée sur liste rouge (Tableau 10) due à leur modèle d'une qualité jugée insuffisante ou incertaine. Ce jugement s'est basé sur le résultat des quatre indices mesurés. Également, pour certaines espèces (Annexe 4 & Tableau 10) ainsi que pour certains modèles (Tableau 8), l'AICc n'est pas renseigné. La raison est que lors des manipulations R, la fonction « `calc.aicc()` » n'est pas parvenu à le mesurer.

Tableau 10: Liste des espèces dont la qualité de leur modèle est jugée insuffisante et donc ne devant pas être pris en compte pour le reste des analyses.

Espèces	Indices							
	AICc	AUC	Mat	PA	UA	Nb data cal	Nb data val	Modèle
<i>Alcedo atthis</i>	1330,774	0,880	0,219	0,875	0,061	73	24	3
<i>Asio otus</i>	9492,146	0,609	0,222	0,769	0,079	471	156	1
<i>Charadrius dubius</i>	/	0,935	0,241	0,929	0,068	44	14	1
<i>Ciconia nigra</i>	/	0,787	0,151	0,667	0,034	83	27	1
<i>Circus aeruginosus</i>	1279,514	0,959	0,231	0,944	0,062	56	18	4
<i>Falco peregrinus</i>	/	0,757	0,162	0,500	0,049	50	16	1
<i>Falco subbuteo</i>	/	0,709	0,159	0,531	0,065	97	32	2
<i>Muscicapa striata</i>	15457,959	0,725	0,357	0,599	0,2	803	267	2
<i>Pernis apivorus</i>	7140,215	0,511	0,206	0,605	0,051	360	119	3
<i>Serinus serinus</i>	2517,757	0,837	0,21	0,667	0,066	111	36	4

Le résultat brut des modélisations présentées en annexe (Annexe 4) montre également via un code couleur les différentes classes de qualité des modèles. Il y en a en tout, quatre différentes allant des plus mauvais modèles comme celui des deux espèces de Falco, celui de la Bondrée apivore et du Hibou moyen-duc aux meilleurs. Au sein de cette dernière catégorie se trouvent le Bruant proyer, la Foulque macroule, la Locustelle tachetée, le Gorgebleue à miroir, la Grèbe huppé et le Tarier des prés. Seule la locustelle tachetée possède un AUC inférieur à 0,9 (0,876) et un PA relativement faible (0,736). Malgré tout, son indice de précision global est l'un des meilleurs (0,456) tout comme son UA (0,295). Parmi les classes intermédiaires, il y a les espèces ayant également été mises sur liste rouge en raison de leur coefficient de Matthew et leur UA trop faibles mais qui possèdent tout de même de bonnes valeurs pour les autres indices. Parmi elles, se trouve le Petit gravelot qui possède un AUC de 0,935 ainsi qu'un PA de 0,929 mais un UA de 0,068 ce qui est assez faible. Finalement, le reste des espèces possèdent des valeurs d'indices relativement bons mais non exceptionnels. Il y a tout de même certains modèles visiblement moins fiables que d'autres qui sont classés dans l'avant dernière classe « neutre ». Parmi eux, il y a celui du Torcol fourmilier avec un PA d'à peine 0,647 et la Chouette effraie avec un AUC de 0,678. D'un point de vue global, 77% des espèces possèdent un AUC supérieur à 0,8 et aucun ne montre de valeurs d'indice impliquant une efficacité plus faible ou égale à un modèle aléatoire.

Pour finir, deux couches rasters sont générées par modèle. Il s'agit en premier lieu de la représentation des « probabilités » d'occurrence de l'espèce offerte par MaxEnt. La valeur de ses pixels informe sur la probabilité de présence de l'espèce modélisée. Il ne s'agit malgré tout pas d'une probabilité classique où une valeur de 0,5 signifierait qu'une espèce ait une chance sur deux d'être présente. La seconde couche représente les points de présence estimés de l'espèce déterminés via un seuil. Chaque pixel ayant une valeur de 1 représente ainsi un habitat complet ou partiel favorable à l'espèce. Ceux ayant une valeur nulle représentent la matrice paysagère. La figure 5 illustre les deux sortes de rasters résultantes de chaque modélisation et ceci pour trois espèces différentes ainsi que les points d'observation fournis par Natagora.

## Résultats

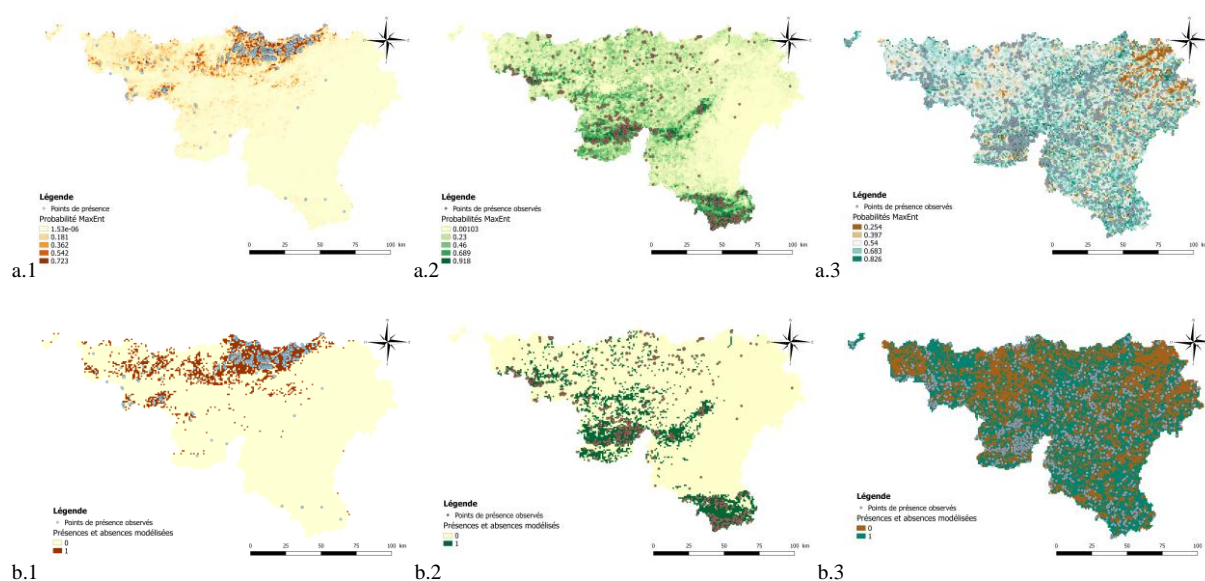


Figure 5: Couche raster représentant les probabilités "MaxEnt" de présence (a) et les présences et absences modélisées (b) de trois espèces différentes choisies selon la qualité de leur modèle respectif. La première est le Bruant proyer (*Emberiza calandra*) avec un modèle jugé très bon. La seconde est le Lorient d'Europe (*Oriolus oriolus*) dont le modèle est jugé comme bon et finalement, la Bondrée apivore (*Bondrus apivorus*) avec un modèle d'une qualité plus faible.

### 3.3. Sélection des espèces indicatrices

Les points de présences obtenus après modélisation qu'il s'agisse des espèces rares ou communes sont séparés selon les quatre régions représentées en figure 6. Le nord du sillon Sambre et Meuse a été regroupé avec le Condroz et le sillon lui-même.

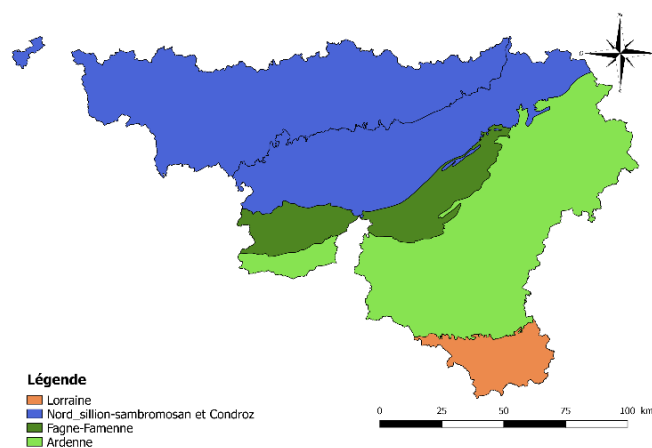
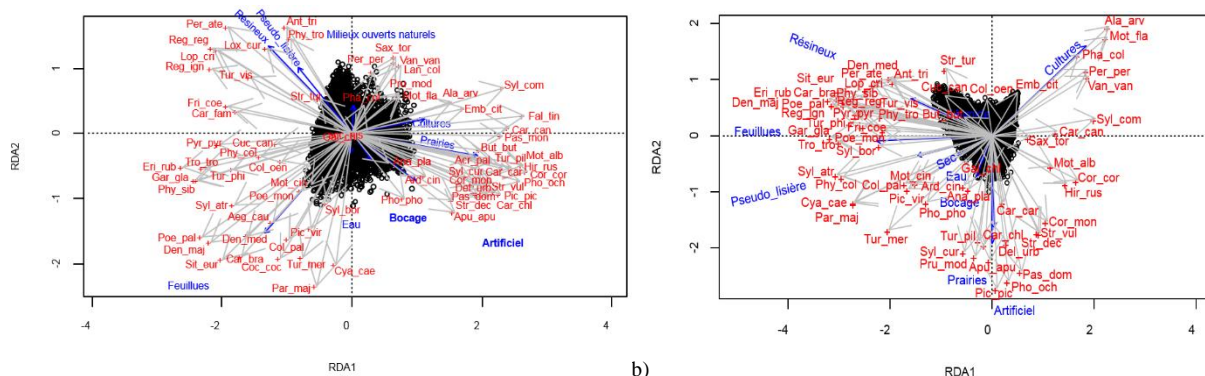


Figure 6: Cartographie des territoires biogéographiques considérés séparément lors de l'analyse des espèces indicatrices.

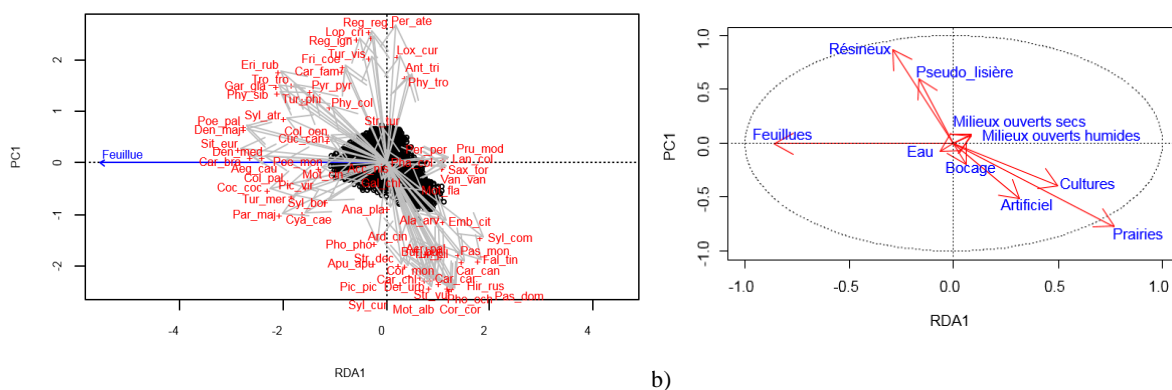
Très vite, la trame des milieux rocheux peut être mise de côté en raison du peu d'informations disponibles dans les couches écotoques à cette échelle. Un réseau centré sur la trame artificielle, n'est également pas construit n'empêchant pas la prise en compte de sa variable écotope (« Artificielle ») dans les analyses qui suivent. Ces dernières doivent être réalisées séparément pour les deux groupes en raison de la nature différente des données offertes par les deux algorithmes de modélisation. Chaque analyse de la redondance donne un graphique représentant l'agencement des espèces selon les différentes variables prises en compte par rapport aux deux axes les plus explicatifs. La figure 7a et b montrent les résultats de l'analyse des espèces communes dans les deux plus grandes régions considérées, l'Ardenne et le nord du sillon/Condroz.

## Résultats



a) b)  
Figure 7: Analyse de la redondance des données de présences prises en Ardenne (a) et au nord du sillon Sambre et Meuse/Condruz (b) avec prise en compte des occupations du sol représentant l'entièreté des trames. Seules les trois premières lettres du genre et de l'espèce sont renseignées.

Une fois l'analyse générale réalisée, une RDA par trame est effectuée afin de déterminer spécifiquement les espèces les plus indicatrices de chaque trame via le  $\cos^2$  qu'ils possèdent par rapport à l'axe de la variable traduisant cette trame. La figure 8 montre le graphique obtenu après l'analyse de la trame feuillue en Ardenne pour les espèces communes. À chaque graphique de ce type, un cercle de corrélation y est systématiquement associé afin de visualiser l'agencement des autres variables d'occupation du sol.



a) b)  
Figure 8: Représentation graphique des résultats de la RDA concernant les données en Ardenne pour la trame "Forêts feuillues" (a) ainsi que le cercle de corrélation associé (b).

Le  $\cos^2$  est la donnée la plus intéressante car elle informe sur le pourcentage de variance de la distribution de l'espèce expliquée par une ou plusieurs variables environnementales. Cependant, les coordonnées ont également toute leur importance dans la caractérisation de la relation entre l'espèce et la trame analysée. En effet, *Sylvia communis* possède un  $\cos^2$  de 39,56% mais n'est pas du tout représentative des forêts feuillues car son axe est opposé à celui de la trame (Fig 8a.). C'est pourquoi seules les espèces ayant un  $\cos^2$  suffisamment élevé et étant positivement corrélée avec la variable en question peuvent être maintenues comme candidate à être indicatrice. Le tableau 11 reprend toutes les espèces rares potentiellement indicatrices de chaque trame en Ardenne avec leur  $\cos^2$ . Ceux des autres trames sont disponibles en annexe 5. Une sélection plus contraignante est ensuite appliquée à ces espèces candidates se basant sur l'écologie de l'espèce, la qualité de son modèle et sa présence ou non dans d'autres trames. Les résultats sont résumés au sein du tableau 12. À chaque tentative, la séparation de la trame des milieux naturels ouverts en milieux secs et humides ne donne pas de résultats suffisants afin d'avoir une espèce représentative des deux trames. La trame humide pose le plus de soucis comme il est observable en figure 9. Les deux trames sont donc regroupées à chaque fois.

## Résultats

Tableau 11: Tableau reprenant les cos<sup>2</sup> de chaque espèce ayant au moins une valeur supérieure au seuil (10%) dans la région des Ardennes et concernant les espèces rares. La valeur « <10 » signifie qu'effectivement le cos<sup>2</sup> est inférieur à 10% ou que les coordonnées de l'espèce sont opposées à celle de la trame.

Nom scientifique	Nom vernaculaire	Trame									
		Agricole	Bocage	Culture	eau	Feuillue	Forêt	Lisière	Ouvert	Prairie	Résineux
Accipiter gentilis	Autour des palombes	<10	<10	<10	<10	<10	10,91	<10	19,451	<10	<10
Acrocephalus scirpaceus	Rousserole effarvatte	<10	<10	<10	12,11	<10	<10	<10	<10	<10	<10
Aegolius funereus	Chouette de Tengmalm	<10	<10	<10	<10	<10	22,423	19,057	<10	<10	32,745
Alcedo atthis	Martin-pêcheur	<10	<10	<10	22,939	<10	<10	<10	<10	<10	<10
Anthus pratensis	Pipit farlouse	<10	<10	<10	<10	<10	<10	<10	32,726	<10	<10
Asio otus	Hibou moyen-duc	10,875	<10	<10	<10	<10	<10	<10	<10	<10	<10
Athene noctua	Chevêche d'Athéna	14,806	16,685	<10	<10	<10	<10	<10	<10	15,045	<10
Carduelis spinus	Tarin des aulnes	<10	<10	<10	<10	<10	<10	12,466	11,741	<10	25,08
Charadrius dubius	Petit gravelot	<10	<10	<10	<10	<10	<10	<10	21,833	<10	<10
Corvus corax	Grand corbeau	<10	<10	<10	<10	<10	<10	<10	12,704	<10	<10
Corvus frugilegus	Corbeau freux	40,307	<10	<10	<10	<10	<10	<10	<10	40,759	<10
Coturnix coturnix	Caille des blés	37,17	<10	49,279	<10	<10	<10	<10	<10	24,282	<10
Cygnus olor	Cygne tuberculé	<10	<10	<10	19,198	<10	<10	<10	<10	<10	<10
Dendrocopos minor	Pic épeichette	<10	<10	<10	<10	45,966	18,648	<10	<10	<10	<10
Dryocopus martius	Pic noir	<10	<10	<10	<10	18,279	52,738	24,174	<10	<10	26,878
Emberiza schoeniclus	Bruant des roseaux	11,91	<10	<10	<10	<10	<10	<10	29,613	12,19	<10
Falco peregrinus	Faucon pèlerin	<10	<10	<10	19,423	<10	<10	<10	<10	<10	<10
Ficedula hypoleuca	Gobemouche noir	<10	<10	<10	<10	57,857	30,783	<10	<10	<10	<10
Fulica atra	Foulque macroule	<10	<10	<10	30,295	<10	<10	<10	<10	<10	<10
Jynx torquilla	Torcol fourmilier	<10	<10	<10	<10	<10	<10	16,809	37,398	<10	<10
Lanius excubitor	Pie-grièche grise	<10	<10	<10	<10	<10	<10	21,107	40,802	<10	<10
Locustella naevia	Locustelle lusciniotide	<10	<10	<10	<10	<10	<10	11,428	32,355	<10	<10
Loxia curvirostra	Bec-croisé des sapins	<10	<10	<10	<10	<10	17,341	21,821	<10	<10	31,222
Lullula arborea	Alouette lulu	<10	<10	<10	<10	<10	<10	51,295	11,006	<10	16,489
Oriolus oriolus	Loriot d'Europe	<10	<10	<10	<10	22,667	<10	<10	<10	<10	<10
Podiceps cristatus	Grèbe huppé	<10	<10	<10	59,491	<10	<10	<10	<10	<10	<10
Riparia riparia	Hirondelle de rivage	<10	<10	<10	27,059	<10	<10	<10	<10	<10	<10
Saxicola rubetra	Grèbe castagneux	<10	<10	<10	<10	<10	<10	<10	32,76	<10	<10
Serinus serinus	Tarier des près	<10	<10	<10	<10	<10	<10	<10	<10	11,594	<10
Strix aluco	Serin cini	<10	<10	<10	<10	19,366	49,817	15,421	<10	<10	22,967
Tachybaptus ruficollis	Chouette hulotte	<10	<10	<10	17,987	<10	<10	<10	<10	<10	<10
Tyto alba	Chouette effraie	46,923	10,051	10,151	<10	<10	<10	<10	<10	47,782	<10

Tableau 12: Tableau récapitulatif reprenant les espèces indicatrices finalement sélectionnées pour chaque trame au sein de chaque région précédemment délimitée. Les espèces communes se situent à droite suivies par les espèces rares.

Région	Trames	Espèces communes/rares							
Ardenne	Milieux agricoles	<i>Carduelis cannabina</i>							
	Prairies	<i>Corvus corone</i>	<i>Phoenicurus ochruros</i>	<i>Motacilla alba</i>	<i>Hirundo rustica</i>	<i>Corvus frugilegus</i>	<i>Tyto alba</i>		
	Cultures	<i>Alauda arvensis</i>	<i>Emberiza citrinella</i>	<i>Coturnix</i>					
	Forêts	<i>Erithacus rubecula</i>	<i>Phylloscopus sibilatrix</i>	<i>Fringilla coelebs</i>	<i>Certhia familiaris</i>	<i>Dryocopus martius</i>	<i>S. aluco</i>		
	Feuillues	<i>Sitta europaea</i>	<i>Poecile palustris</i>	<i>Dendrocopos major</i>	<i>Dendrocopos medius</i>	<i>Ficedula hypoleuca</i>	<i>Dendrocopos minor</i>		
	Résineux	<i>Periparus ater</i>	<i>Regulus</i>	<i>Lophophanes cristatus</i>	<i>Aegolius funereus</i>	<i>Laxicola curvirostra</i>	<i>Carduelis spinus</i>		
	Milieux ouverts	<i>Saxicola rubetra</i>	<i>Jynx torquilla</i>	<i>Anthus pratensis</i>					
	Lisières	<i>Anthus trivialis</i>	<i>Phylloscopus trochilus</i>	<i>Lullula arborea</i>					
	Bocages	<i>Athene noctua</i>							
	Eaux	<i>Anas platyrhynchos</i>	<i>Gallinula chloropus</i>	<i>Prodicops cristatus</i>	<i>Fulica atra</i>	<i>Riparia riparia</i>			
Nord sillon Sambre et Meuse et Condroz	Milieux agricoles								
	Bocages	<i>Sylvia curruca</i>	<i>Turdus pilaris</i>						
	Prairies	<i>Athene noctua</i>	<i>Tyto alba</i>						
	Cultures	<i>Alauda arvensis</i>	<i>Motacilla flava</i>	<i>Perdix</i>	<i>Corurnix coturnix</i>	<i>Emberiza calandra</i>			
	Forêts	<i>Erithacus rubecula</i>	<i>Dendrocopos major</i>	<i>Garrulus glandarius</i>	<i>Dryocopus martius</i>				
	Feuillues	<i>Phylloscopus collybita</i>	<i>Dendrocopos minor</i>	<i>Strix aluco</i>	<i>Accipiter gentilis</i>				
	Résineux	<i>Periparus ater</i>	<i>Lophophanes cristatus</i>	<i>Laxicola curvirostra</i>					
	Milieux ouverts	<i>Jynx torquilla</i>							
	Lisières	<i>Phylloscopus trochilus</i>	<i>Phylloscopus collybita</i>	<i>Garrulus glandarius</i>	<i>Locustella naevia</i>	<i>Lullula arborea</i>	<i>Muscicapa striata</i>		
	Eaux	<i>Anas platyrhynchos</i>	<i>Prodicops cristatus</i>	<i>Riparia</i>	<i>Fulica atra</i>				
Lorraine	Prairies	<i>Carduelis</i>	<i>Hirundo rustica</i>	<i>Sylvia communis</i>	<i>Tyto alba</i>	<i>Corvus frugilegus</i>			
	Cultures	<i>Alauda arvensis</i>	<i>Emberiza citrinella</i>						
	Forêts	<i>Dendrocopos major</i>	<i>Erithacus rubecula</i>	<i>Garrulus glandarius</i>	<i>Dryocopus martius</i>	<i>Strix aluco</i>	<i>Laxicola curvirostra</i>		
	Feuillues	<i>Sitta europaea</i>	<i>Carthia brachydactyla</i>	<i>Poecile palustris</i>					
	Résineux	<i>Regulus</i>	<i>Periparus ater</i>						
	Eaux	<i>Anas platyrhynchos</i>	<i>Prodicops cristatus</i>	<i>Fulica atra</i>	<i>Tachybaptus ruficollis</i>				
	Lisières	<i>Anthus trivialis</i>	<i>Phylloscopus trochilus</i>	<i>Phylloscopus collybita</i>	<i>Lullula arborea</i>				
	Milieux ouverts	<i>Jynx torquilla</i>	<i>Emberiza schoeniclus</i>	<i>Acrocephalus scirpaceus</i>					
Fagne-Famenne	Milieux agricoles	<i>Sylvia communis</i>							
	Cultures	<i>Alauda arvensis</i>	<i>Emberiza citrinella</i>	<i>Corurnix coturnix</i>	<i>F. subbuteo</i>				
	Prairies	<i>Carduelis</i>	<i>Hirundo rustica</i>	<i>Phoenicurus ochruros</i>	<i>Athene noctua</i>	<i>Tyto alba</i>	<i>Corvus frugilegus</i>		
	Bocages								
	Forêts	<i>Dendrocopos major</i>	<i>Erithacus rubecula</i>						
	Feuillues	<i>Poecile palustris</i>	<i>Sitta europaea</i>	<i>Dendrocopos medius</i>	<i>Ficedula hypoleuca</i>	<i>Dendrocopos minor</i>	<i>Dryocopus martius</i>		
	Résineux	<i>Periparus ater</i>	<i>Regulus</i>	<i>Laxicola curvirostra</i>	<i>Corvus corax</i>	<i>Dryocopus martius</i>			
	Eaux	<i>Prodicops cristatus</i>	<i>L. svecica</i>	<i>Tachybaptus ruficollis</i>					
	Lisières	<i>Phylloscopus collybita</i>	<i>Phylloscopus trochilus</i>	<i>Anthus trivialis</i>	<i>Dryocopus martius</i>	<i>Lullula arborea</i>	<i>Dryocopus martius</i>		
	Milieux ouverts	<i>Jynx torquilla</i>	<i>Lullula arborea</i>	<i>Locustella naevia</i>					





## Résultats

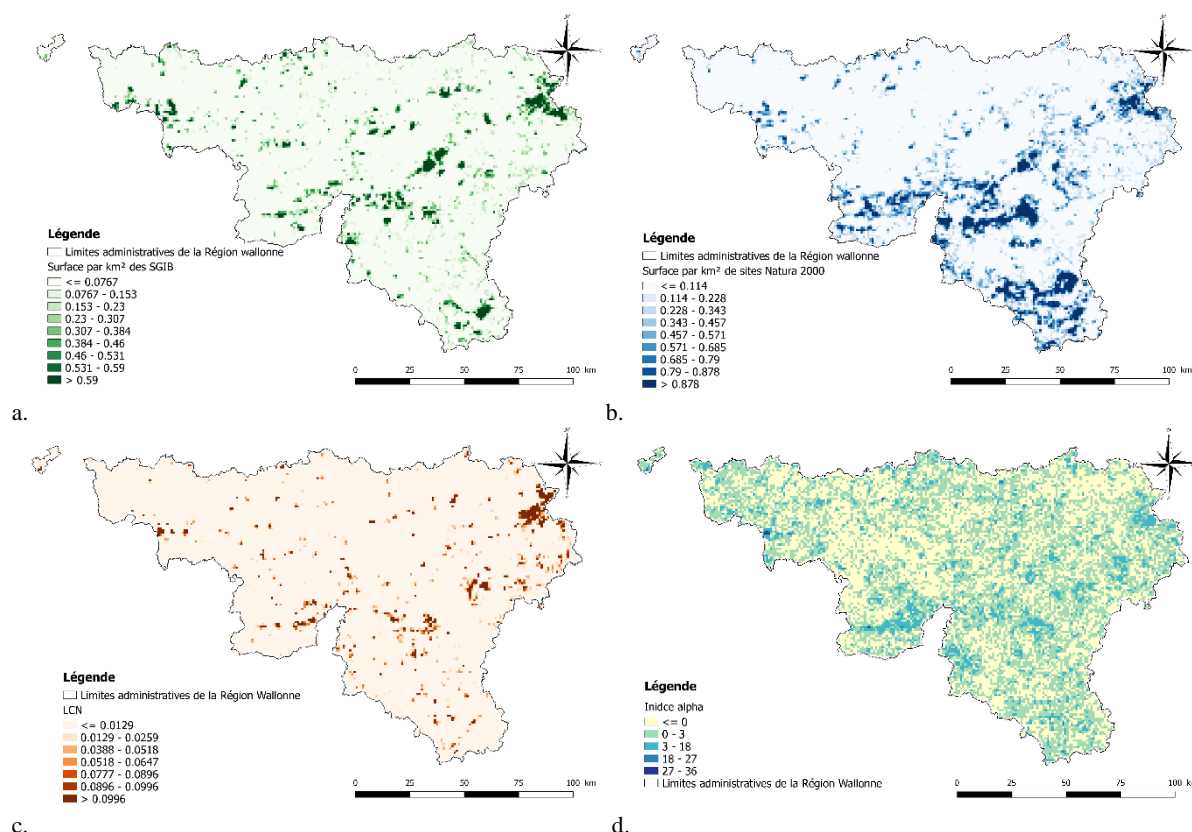


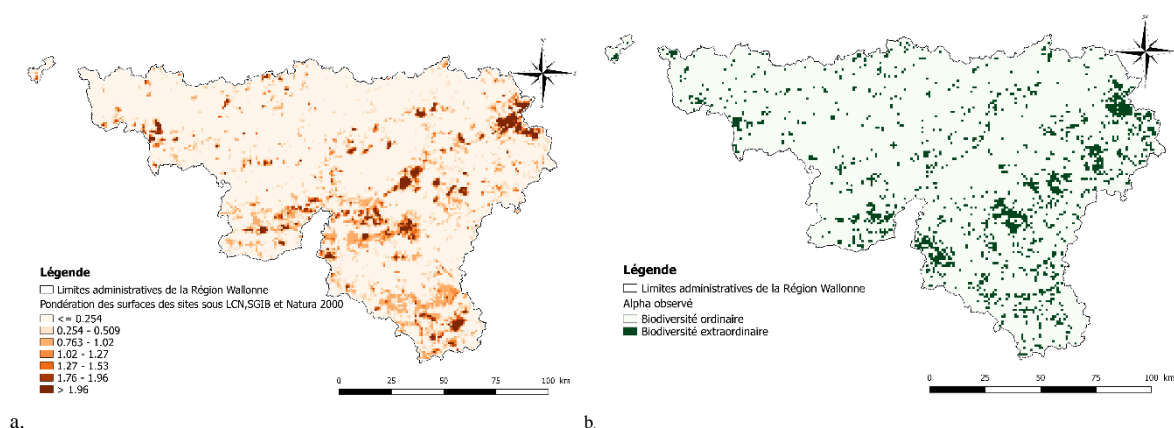
Figure 10: Cartes des proportions surfaciques au sein d'un km<sup>2</sup> des différents statuts relatif à la conservation de la nature en Belgique. La figure a. représente les SGIB, la b. affiche les sites Natura 2k alors que la c. montre les sites sous protection de la LCN. La figure d représente quant à elle, l'indice alpha observé.

Les trois premières cartes sont additionnées l'une à l'autre avec une pondération favorisant la LCN et les SGIB (Figure 11a). Ensuite, un seuil est déterminé pour cette carte résultante ainsi que celle de l'indice alpha observé (Tableau 13). Pour les sites protégés ou d'intérêt biologique, la classification binaire de Jenks donne un seuil égal à 1. La figure 11 représente les deux cartes permettant la délimitation du réseau écologique existant (Figure 12) qui est représenté par la carte résultante de la somme des surfaces sous statut (a.) et celle de l'indice alpha observé sous forme binaire (b.).

Tableau 13: Nombre d'espèces seuil différenciant des autres, un carré kilométrique ayant une biodiversité observée remarquable.

Région biogéographique	Seuil (nbr d'espèces)
Ardenne	2
Condroz	3
Fagne	6
Lorraine	3
NSSM	4

## Résultats



a.

b.

Figure 11: Représentation cartographique de la somme des surfaces par carré kilométrique pondérées des sites sous statut (a.) ainsi que le résultat de l'application des seuils sur l'indice alpha observé (b.).

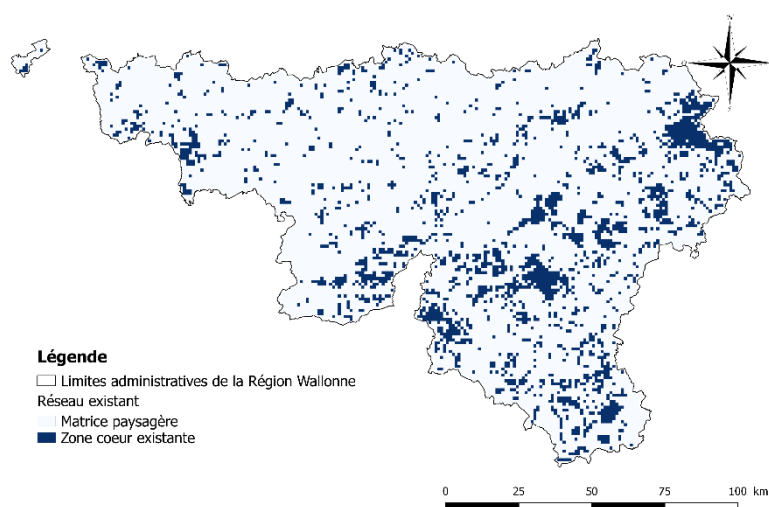


Figure 12: Ensemble des zones cœurs existantes selon la diversité observée et les statuts de protection ou autres dénominations relatives à la conservation de la biodiversité.

Pour les analyses qui vont suivre, il est préférable d'également définir des seuils de différenciation des carrés kilométriques ayant suffisamment de surface sous LCN ou de SGIB pour qu'il soit entièrement considéré sous le statut correspondant (Tableau 14). Il en est de même concernant le réseau Natura 2000.

Tableau 14: Seuils appliqués aux cartes rasters afin de déterminer sous quel statut de protection un carré kilométrique peut être assimilé.

Statut de protection	Seuil
SGIB	0,3
Natura 2000	0,36
LCN	0,3



### 3.5. Réseau écologique potentiel

Dans le but de réaliser le réseau écologique potentiel qui pour rappel, se base uniquement sur les données modélisées, deux cartes sont nécessaires ainsi que le nombre d'espèces indicatrices par trame au sein de chaque carré kilométrique de Wallonie. La première carte informe sur la diversité modélisée d'un site via l'indice de diversité alpha (Figure 13a). La seconde correspond à ce même indice alpha mais pondéré avec les résultats des RDA du point 3.3, uniquement pour les espèces rares et pour chaque trame. La Figure 13b. représente les résultats pour la trame des forêts feuillues. Lorsque seule une présence d'une espèce indicatrice et la valeur de l'indice alpha modélisé sont prises en compte pour la détermination des réseaux, cela une très faible précision des réseaux. En effet, ils se répandent sur une très grande proportion de la Wallonie et les trames écologiques ne sont plus représentées par leur réseau. Il est donc nécessaire de tester une méthode alternative qui dans ce cas-ci, consiste à pondérer l'indice alpha.

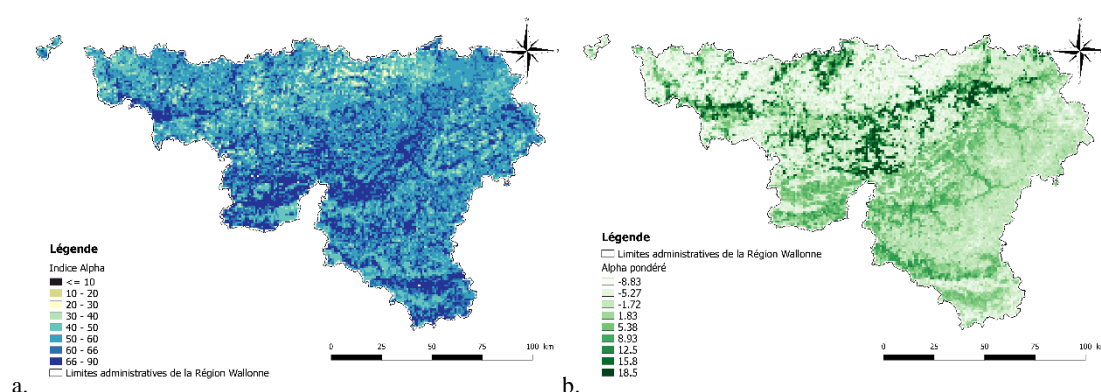


Figure 13: Couche raster représentant l'indice alpha mesuré à l'aide des présences modélisées de toutes les espèces (a.) et celui pondéré par le résultat des RDA pour la trame forestière (b.).

Concrètement, un seuil de différenciation des carrés kilométriques ayant une bonne biodiversité modélisée est déterminé. Comme précédemment, cela doit être fait par région biogéographique. Comme les valeurs de l'indice par région suivent une distribution plus ou moins normale, c'est la moyenne qui est choisie comme seuil. Ainsi, la classification de Jenks n'est utilisée que pour l'indice pondéré et également séparément par région. Les seuils résultants de cette classification se trouvent en tableau 15. Ceux relatifs de l'indice classique sont dans le tableau 16.

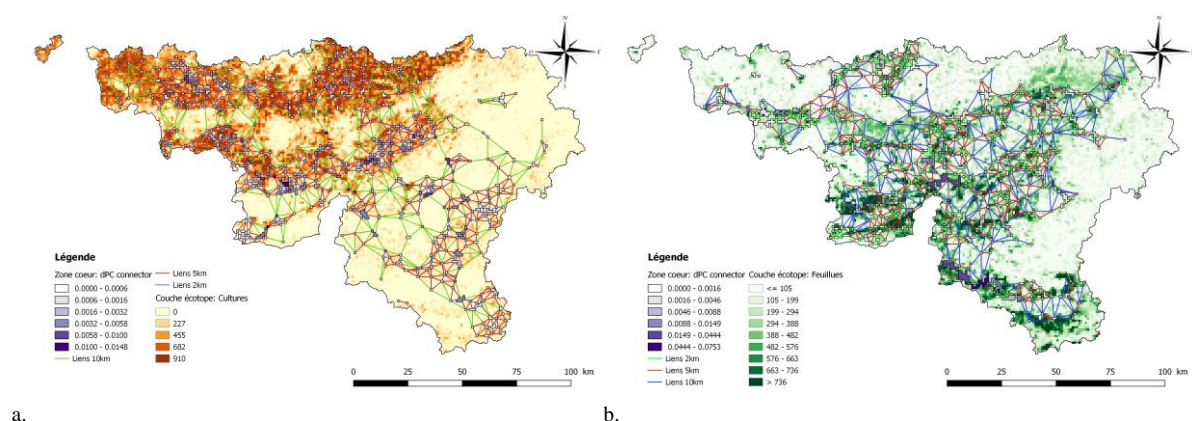
Tableau 15: Seuils de différenciation imposés à la valeur de l'indice alpha pondéré par carré kilométrique par région et pour chaque trame écologique.

Région	Trame écologique									
	Milieux agricoles	Bocages	Cultures	Milieux aquatiques	Forêts décidues	Forêts	Lisières	Milieux ouverts	Prairies	Forêts résineuses
Ardenne	0,339	0,898	0	9,806	6,572	9,751	11,379	15,528	0,446	10,293
Condroz	1,455	0,875	1,772	18,658	16,145	16,588	14,490	13,900	4,874	8,652
Fagne	3,096	2,422	0	6,109	6,205	6,347	4,000	10,093	3,349	5,216
Lorraine	3,101	2,902	0,024	9,995	4,030	4,166	4,128	12,683	4,965	3,506
NSSM	5,669	0,376	3,285	21,782	11,392	11,548	13,617	16,956	4,429	4,195

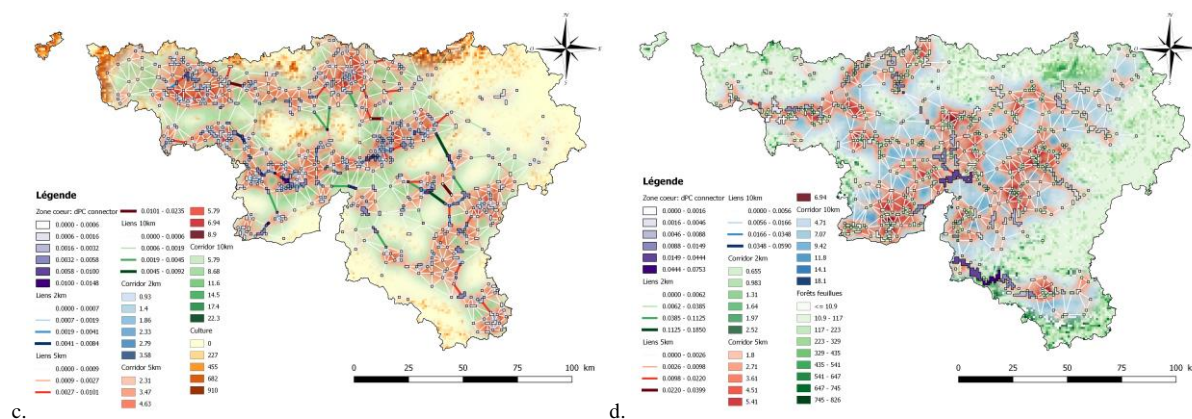
Tableau 16: Seuils du nombre d'espèces imposés aux carrés kilométriques pour leur sélection en zone cœur des cinq régions.

Région biogéographique	Seuil (nbr d'espèces)
Ardenne	57
Condroz	59
Fagne	66
Lorraine	63
NSSM	53

De cette manière, un carré kilométrique désigné comme zone cœur doit dans un premier temps posséder une espèce indicatrice de la trame en question. Dans un second temps, il doit respecter les deux critères de richesse spécifique imposés d'un côté par l'indice alpha modélisé et de l'autre, par cet indice pondéré. Le résultat est une série de couches rasters binaires dont les pixels ayant une valeur de 1 sont considérés comme des zones cœurs de biodiversité de la trame correspondante. Ces derniers sont chacun à leurs tours, manipulés dans Graphab afin de déterminer les différents liens existants selon les trois distances de dispersion testées. La figure 14 représente le résultat final après traitement et calcul des dPC pour chaque distance de la trame des cultures (a) et des forêts feuillues (b). Il y a donc quatre shapefiles résultants dont l'un contient toutes les zones centrales de la trame et les trois autres contiennent les liens entre elles pour chaque distance de dispersion.



a. b.  
Figure 14: Ensemble des liens pouvant exister entre les différentes taches d'habitat favorable selon 3 distances de dispersion moyennes et ceci, pour la trame des cultures (a.) et des forêts feuillues (b.). Il est également à noter que le dPC connector affiché est mesuré via la distance de dispersion max de 10km.



c. d.  
Figure 15: Représentation cartographique du réseau écologique final de la trame des cultures (a.) et des forêts feuillues (b.) où les liens directs et fictifs entre les taches sont représentés par des lignes et les zones de connexions pouvant devenir des corridors sont représentées sous forme de carte de chaleur. Également, le dPC connector affiché est relatif au réseau construit via une distance de dispersion de 10km.

La figure 16 peut directement être comparée avec la précédente et met en évidence pour les mêmes trames, les liens n'ayant pas d'utilité fonctionnelle et qui sont par conséquent, retirés. Elle représente ainsi le réseau écologique final de deux manières différentes. La première est linéaire, se composant uniquement des lignes unidimensionnelles liant le centre d'une zone cœur à celui d'une autre. Ensuite arrive la représentation bidimensionnelle des corridors où l'aspect d'épaisseur est enfin pris en compte. Cette dernière facilite grandement la visualisation plus des milieux pouvant être utilisés pour le déplacement des espèces du réseau. Dans la continuité de la cartographie des réseaux écologiques, la figure 16 reprend les cartes du réseau final de la trame des forêts résineuses et des prairies (a et b) de la même manière que la figure 15 et deux autres cartes.

Ces dernières possèdent des couches de chaleur pondérées qui montrent plus précisément les zones les plus importantes dans la connectivité au sein de l'entière du réseau écologique considéré (Figure 16c et d).

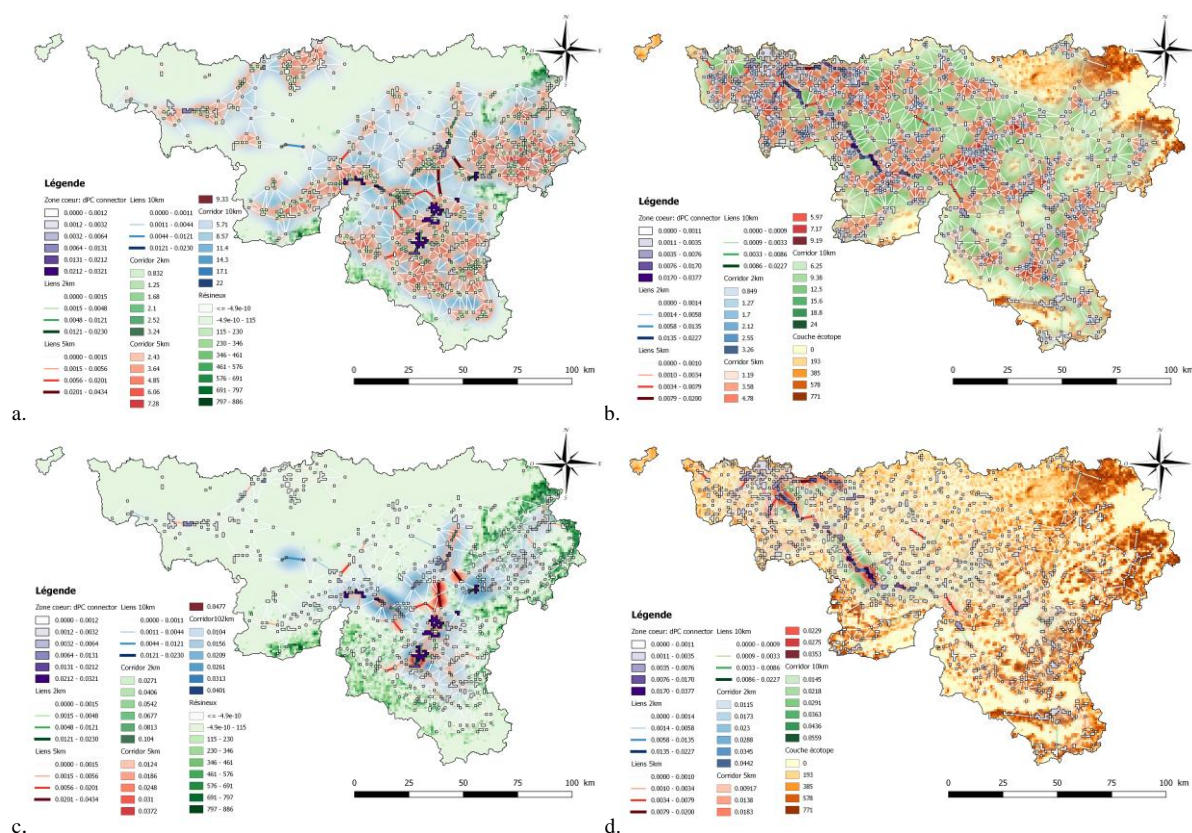


Figure 16: Comparaison cartographique des corridors représentés via une carte de chaleur avec et sans pondération selon le dPC connector des liens. La carte a et c représentent la trame des forêts résineuses et les deux autres représentent la trame des prairies.

Deux indices globaux sont mesurés pour l'entière des trames écologiques étudiées afin de caractériser au mieux la connectivité et la qualité de chaque réseau (Tableau 17). Également, les valeurs moyennes de ces deux indices ainsi que leur écart-type par distance de dispersion sont disponibles dans le tableau 18. Concernant le delta métrique dPC et ses trois composantes, ils sont incorporés au sein de la table d'attribut des shapefiles représentant les zones cœurs de biodiversité et ceci pour chaque trame écologique. La représentation cartographique du dPC connector est pour rappel, disponible sur toutes les figures affichant les zones cœurs d'un réseau et sur les cartes en annexe 6.

## Résultats

Tableau 17: Résultats des indices de probabilité de connexion et d'Harary pour les réseaux de chaque trame ainsi que pour les trois distances de dispersion.

Trames écologiques	d (km)	Indices globaux		Harary		nbr ZC	Surface (km²)
		PC	Harrary	Valeur max	Valeur relative		
Forêts feuillues	10	1,394E-04	15089,801	124251	12,145	499	1042
	5	5,469E-05	8915,660	124251	7,176	499	1042
	2	2,165E-05	1218,171	124251	0,980	499	1042
Forêts résineuses	10	9,002E-05	16433,325	142845	11,504	535	864
	5	3,470E-05	9274,021	142845	6,492	535	864
	2	1,044E-05	1438,137	142845	1,007	535	864
Forêts	10	1,205E-04	16838,819	138075	12,195	526	952
	5	4,539E-05	9089,357	138075	6,583	526	952
	2	1,298E-05	1250,825	138075	0,906	526	952
Cultures	10	6,719E-05	19528,966	180901	10,795	602	851
	5	2,212E-05	8794,545	180901	4,862	602	851
	2	5,769E-06	1445,846	180901	0,799	602	851
Prairies	10	2,217E-04	36016,561	359128	10,029	848	1397
	5	8,766E-05	26645,417	359128	7,419	848	1397
	2	2,231E-05	4853,744	359128	1,352	848	1397
Milieux a gricoles	10	1,241E-04	22945,858	201930	11,363	636	956
	5	4,700E-05	13792,664	201930	6,830	636	956
	2	1,160E-05	2658,270	201930	1,316	636	956
Milieux ouverts	10	1,539E-05	1276,313	11026	11,575	149	328
	5	1,030E-05	644,786	11026	5,848	149	328
	2	7,002E-06	300,936	11026	2,729	149	328
Lisière	10	1,022E-04	14089,276	123256	11,431	497	967
	5	3,994E-05	6481,281	123256	5,258	497	967
	2	1,519E-05	1070,846	123256	0,869	497	967
Bocages	10	5,707E-05	12724,169	98790	12,880	445	691
	5	2,028E-05	5770,301	98790	5,841	445	691
	2	5,018E-06	916,729	98790	0,928	445	691
Milieux aquatiques	10	1,180E-04	7316,605	68635	10,660	371	912
	5	5,282E-05	2067,505	68635	3,012	371	912
	2	3,187E-05	823,553	68635	1,200	371	912

Tableau 18: Valeurs moyennes des indices par distance de dispersion ainsi que leur écart-type.

	d(km)	PC	Harary relatif
Moyenne	10	1,056E-04	11,458
	5	4,149E-05	5,932
	2	1,438E-05	1,209
Ecart-type	10	3,919E-05	0,602
	5	1,602E-05	0,968
	2	6,698E-06	0,354



Dans tous les cas, lorsque les résultats des deux indices sont comparés chez la même espèce, au plus la distance de dispersion est grande au plus les valeurs sont élevées. Leur différence informe sur la qualité perdue lors de la diminution de la distance et inversement. Par exemple, la réduction de la distance de dispersion à 5km fait diminuer la PC du réseau des milieux ouverts de  $1,539 \cdot 10^{-5}$  à  $1,03 \cdot 10^{-5}$ . Le gain en PC le plus important concerne la trame des prairies lorsque la distance de dispersion passe de 5km ( $PC = 8,766 \cdot 10^{-5}$ ) à 10km ( $PC = 2,217 \cdot 10^{-4}$ ). Par contre, l'indice d'Harary révèle la plus grande différence entre la distance de 2 et 5km de cette même trame. Le plus souvent, celle de la PC entre ces deux distances est moins grande que celle entre 5km et 10km. Ça n'est cependant pas le cas pour l'autre indice où elles sont beaucoup plus variables. En comparant les réseaux entre eux désormais, le réseau des prairies possède la plus grande PC avec la valeur de  $2,217 \cdot 10^{-4}$  ainsi que les plus grandes valeurs de l'autre indice pour la distance de 10km. A contrario, La valeur de la PC la plus faible correspond aux milieux bocagers avec une valeur de  $5,018 \cdot 10^{-6}$ . Celle de l'indice d'Harary apparaît chez le réseau des milieux ouverts.

La trame des prairies, des forêts feuillues et mixtes, des milieux agricoles et aquatiques ont pour chaque distance une valeur de PC supérieure à la moyenne sauf pour la distance de 2km où les lisières remplacent les milieux agricoles et forestiers. Également, la trame des prairies est le plus souvent première. Cet indice a l'avantage d'être borné de la même façon, peu importe les caractéristiques constitutives des réseaux ce qui n'est pas le cas pour Harry. Après mesure de la valeur relative de celui-ci pour chaque distance, les bocages obtiennent le meilleur pourcentage pour la plus grande distance. Le classement des trames par rapport à cet indice est bien plus variable que la PC. Par exemple, la trame des prairies est classée dernière avec une valeur de 10,03% pour la distance maximale et passe première avec une valeur de 7,42% pour la distance intermédiaire. Également, les milieux ouverts se classent assez bien pour les distances extrêmes mais pas pour la distance de 5km. Cette variabilité apparaît également dans la liste des trames ayant une valeur supérieure à la moyenne. Concernant les surfaces totales de chaque réseau, la trame des prairies enregistre la plus grande superficie avec 1 397km<sup>2</sup> de zones cœurs de biodiversité. À l'opposé se trouve la trame des milieux naturels ouverts avec une surface de seulement 328km<sup>2</sup> alors que la deuxième surface la plus faible est de 691km<sup>2</sup> pour la trame bocagère.

En conclusion de la cartographie des réseaux potentiels, un total de trois cartes est généré par trame écologique ce qui donne un total de 30 cartes toutes disponibles en annexe (Annexe 6). En effet, une toute dernière carte est ajoutée par trame. Celle-ci représente les carrés kilométriques des réseaux qui sont déjà reconnus via les SGIB ou sont déjà sous protection stricte grâce à la LCN.

### 3.6. Analyses supplémentaires

Le tableau 19 est à double entrée représentant les surfaces partagées entre les réseaux de deux trames différentes. Ainsi, chaque valeur est égale à la somme de la surface de la trame présentée en colonne partagée avec celle citée en ligne divisée par la surface totale de la première.

Tableau 19: Tableau reprenant la proportion de surface partagée entre deux réseaux. Le sens de lecture du tableau à deux entrées est indiqué par la flèche.

Trame comparée Trame de comparaison	Agricole	Bocage	Culture	Eau	Feuillue	Forêt	Lisière	Ouvert	Prairie	Résineux
Agricole	1,000	0,189	0,600	0,009	0,000	0,000	0,000	0,003	0,530	0,000
Bocage	0,262	1,000	0,135	0,029	0,114	0,077	0,045	0,000	0,543	0,038
Culture	0,675	0,109	1,000	0,004	0,000	0,000	0,000	0,000	0,336	0,000
Eau	0,010	0,022	0,003	1,000	0,424	0,280	0,345	0,073	0,080	0,114
Feuillue	0,000	0,076	0,000	0,371	1,000	0,743	0,484	0,010	0,007	0,346
Forêt	0,000	0,056	0,000	0,268	0,813	1,000	0,576	0,032	0,001	0,533
Lisière	0,000	0,032	0,000	0,326	0,521	0,567	1,000	0,189	0,014	0,560
Ouvert	0,009	0,000	0,000	0,204	0,030	0,091	0,558	1,000	0,073	0,375
Prairie	0,363	0,268	0,205	0,052	0,005	0,001	0,010	0,017	1,000	0,000
Résineux	0,000	0,030	0,000	0,120	0,418	0,587	0,627	0,142	0,000	1,000

La proportion la plus élevée mesurée concerne les forêts « mixtes » avec les forêts feuillues enregistrant la valeur de 81,3%. Celle-ci tombe à 53,3% avec les forêts résineuses. Les milieux de culture partagent peu de zones cœurs avec les prairies à raison de 33,6% et 20,5% lorsque la comparaison est inversée. Les trois relatives aux forêts et celle des lisières ne partagent aucune zone cœurs avec les milieux agricoles et très peu avec la sous-trame des prairies. Également, les milieux agricoles ne partagent que 9% de sa surface avec la trame des milieux aquatiques et les valeurs de ses sous-trames sont également basses. Celle des forêts résineuses montre également une faible proportion avec les milieux aquatiques. D'un point de vue général, les milieux ouverts partagent peu de leur surface mis à part avec les forêts résineuses (37,5%) et les lisières (58,7%).

Le nombre de fois qu'un carré kilométrique a été désigné comme zone centrale est calculé et les résultats sont disponibles au sein de la figure 17. La valeur maximale obtenue est cinq trames communes pour 24 carrés kilométriques. Comme il peut être observé sur le graphique (Figure 17b), il y a de nombreuses zones cœurs communes à deux et trois trames. Le nombre diminue fortement à partir de quatre. Malgré tout, un peu plus de 57% de celles-ci sont partagées entre au moins deux trames. La proportion de surface de la Région Wallonne recouverte par l'entière des réseaux écologiques fonctionnels est de 25,62%.

## Résultats

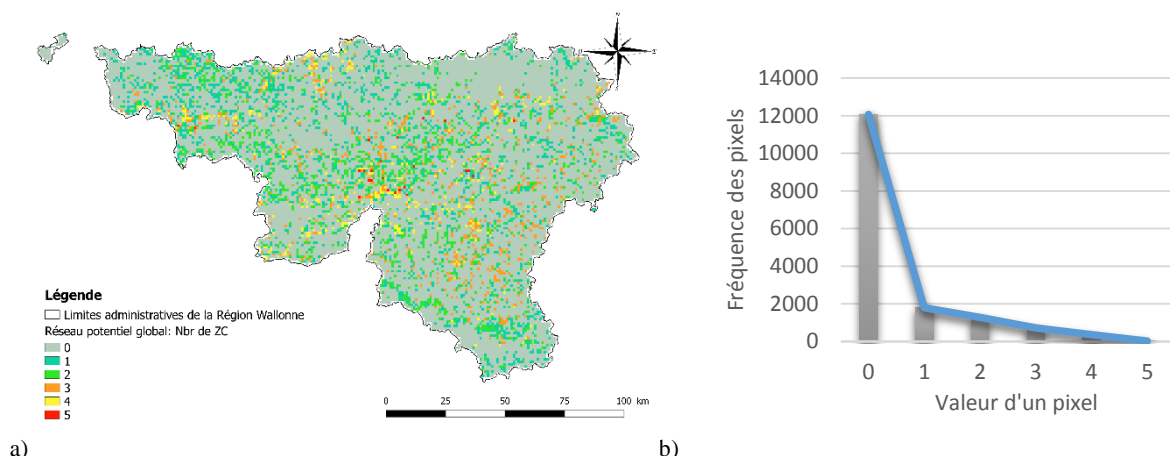


Figure 17: Carte représentant les carrés kilométriques repris en tant que zone cœur ainsi que la fréquence correspondante (a) et le graphique représentant la fréquence des valeurs des pixels de cette couche raster (b).

Afin d'évaluer la représentativité des trames écologiques au sein de leur réseau respectif, le tableau 20 reprend dans l'ordre et pour chaque région, les surfaces occupées par la trame prises dans le réseau, la surface totale de la trame et la fraction de celle-ci comprise dans le réseau. Également, il contient la proportion d'occupation du sol d'une trame au sein de son réseau. Celle-ci peut être interprétée comme la précision d'un réseau. Pour finir, les différences de surfaces moyennes de chaque trame en dehors et au sein des zones cœurs d'un réseau exprimées par rapport à la moyenne hors des zones sont disponibles dans la dernière colonne.

Il semblerait que pour toutes les régions, le réseau des milieux agricoles. Il est donc le réseau le plus spécifique à sa trame. Cependant, il n'est pas représenté en Lorraine et à sa place se trouvent les prairies, les forêts mixtes et feuillues avec une valeur aux alentours des 50%. Pour ce qui est de la fraction de la trame protégée, généralement, la trame des milieux aquatiques prend la tête du classement suivi des milieux ouverts. Dans le Condroz, cependant, cette dernière possède une très faible fraction car seulement une zone cœur de cette trame y est présente. Également, les milieux forestiers au sens large possèdent une très faible superficie à l'intérieur de leur réseau en Lorraine. Concernant plus précisément les milieux aquatiques, même si dans chaque région, plus de 50% des milieux aquatiques sont couverts par son réseau, celui-ci possède généralement la plus faible précision. Or, après analyse de la différence relative des précisions hors et au sein des zones cœurs, elle passe dans les premières suivie des milieux ouverts. Les prairies et les milieux forestiers autres que résineux possèdent d'assez bons résultats pour chaque indice. À deux reprises, la trame bocagère affiche un gain de précision négatif de 33,36 et 85,02% respectivement dans le Condroz et au nord du sillon alors que toutes les autres sont positives.

## Résultats

Tableau 20: Tableaux regroupant les surfaces d'occupation du sol par trame écologique et par région ainsi que la surface de celles-ci sous protection du réseau écologique correspondant, la surface totale de chaque réseau, la précision de celui-ci et le gain en surface moyenne en km<sup>2</sup>. Le tableau a regroupe l'Ardenne et le Condroz, le b concerne la Fagne-Famenne et la Lorraine alors que la c, le Nord du sillon Sambre et Meuse.

Trame	Ardenne					Condroz				
	Surface protégée (km <sup>2</sup> )	Surface totale (km <sup>2</sup> )	Fraction protégée (%)	Surface totale du réseau (km <sup>2</sup> )	Précision du réseau (%)	Surface moyenne Hors ZC (km <sup>2</sup> )	Surface moyenne ZC (km <sup>2</sup> )	Surface moyenne Hors ZC (km <sup>2</sup> )	Surface moyenne ZC (km <sup>2</sup> )	Gain (%)
Milieux agricoles	245,512	1,885,617	13,020	337,000	72,852	0,343	0,729	112,535	288,376	1933,955
Bocages	27,541	184,572	14,921	117,000	23,539	0,147	0,235	59,672	39,500	352,392
Cultures	35,703	256,928	13,896	233,000	15,323	0,064	0,153	141,303	93,141	929,690
Milieux aquatiques	12,992	22,629	57,413	270,000	4,812	0,005	0,048	807,977	17,853	33,351
Forêts décidées	175,722	1,325,908	13,253	356,000	49,360	0,236	0,494	108,920	153,909	1064,931
Forêts mixtes	214,383	2,795,540	7,669	306,000	70,060	0,098	0,701	40,644	149,855	1154,044
Milieux ouverts	45,809	157,242	29,133	224,000	20,540	0,030	0,205	570,575	0,122	24,917
Prairies	217,152	1,628,689	13,333	358,000	60,657	0,296	0,607	104,836	101,036	1004,265
Forêts résineuses	144,493	1,469,632	9,832	438,000	32,989	0,262	0,330	25,929	6,244	89,113
										7,007
										105,000
										0,059
										0,026
										132,160

Trame	Fagne-Famenne					Lorraine				
	Surface protégée (km <sup>2</sup> )	Surface totale (km <sup>2</sup> )	Fraction protégée (%)	Surface totale du réseau (km <sup>2</sup> )	Précision du réseau (%)	Surface moyenne Hors ZC (km <sup>2</sup> )	Surface moyenne ZC (km <sup>2</sup> )	Surface moyenne Hors ZC (km <sup>2</sup> )	Surface moyenne ZC (km <sup>2</sup> )	Gain (%)
Milieux agricoles	54,632	727,190	7,513	68,000	80,341	0,458	0,803	75,555	0	376,858
Bocages	27,053	236,886	11,420	81,000	33,399	0,239	0,334	39,722	0	12,570
Cultures	28,433	201,328	14,123	82,000	34,674	0,136	0,347	154,037	8,300	75,193
Milieux aquatiques	8,474	12,901	65,685	81,000	10,462	0,010	0,105	969,609	1,555	2,452
Forêts décidées	47,552	618,672	7,686	92,000	51,687	0,387	0,517	33,505	5,034	294,140
Forêts mixtes	83,910	709,303	11,830	144,000	58,271	0,444	0,583	31,279	8,398	342,884
Milieux ouverts	8,367	32,127	26,044	27,000	30,989	0,023	0,310	1241,723	9,144	19,933
Prairies	55,947	525,862	10,639	106,000	52,780	0,331	0,528	59,486	43,277	301,665
Forêts résineuses	12,623	90,631	13,928	141,000	8,952	0,057	0,090	57,751	2,607	48,744
										5,348
										19,000
										0,059
										0,137
										131,950

Trame	Nord-Sillon Sambre et Meuse				
	Surface protégée (km <sup>2</sup> )	Surface totale (km <sup>2</sup> )	Fraction protégée (%)	Surface totale du réseau (km <sup>2</sup> )	Précision du réseau (%)
Milieux agricoles	149,538	3601,255	4,152	177,000	84,485
Bocages	1,136	51,025	2,227	218,000	0,521
Cultures	206,685	2532,396	8,162	318,000	64,995
Milieux aquatiques	17,420	33,322	52,278	256,000	6,805
Forêts décidées	135,407	898,569	15,069	286,000	47,345
Forêts mixtes	122,700	935,659	13,114	225,000	54,533
Milieux ouverts	0,810	31,182	2,598	6,000	13,500
Prairies	197,702	1068,859	18,497	595,000	33,227
Forêts résineuses	8,289	37,090	22,348	161,000	5,148
					0,008
					0,051
					568,229



Le tableau 21 représente les valeurs moyennes de l'indice alpha observé et modélisé par carré kilométrique par réseau écologique et par région biogéographique.

Tableau 21: Tableaux récapitulatifs des moyennes ainsi que de leurs écarts-types des indices alpha observé et modélisé par carré kilométrique et par trame écologique pour le Condroz et l'Ardenne (a), la Fagne-Famenne et la Lorraine (b) et le Nord du Sillon Sambre et Meuse (c).

Trame	Ardenne				Condroz			
	Alpha observé	Ecart-type	Alpha model	Ecart-type	Alpha observé	Ecart-type	Alpha model	Ecart-type
Milieux agricoles	1,178	1,750	63,593	4,677	0,674	1,091	62,591	2,768
Bocages	1,675	2,235	68,120	6,498	1,091	1,631	66,240	4,636
Cultures	1,155	1,794	62,948	4,246	0,709	1,396	62,564	2,684
Milieux aquatiques	2,807	2,659	70,937	7,547	2,534	3,787	68,041	6,828
Forêts feuillues	2,000	2,152	67,171	6,463	1,742	2,593	69,054	5,393
Forêts	2,376	2,695	64,549	5,013	1,722	2,662	68,859	5,371
Lisières	3,555	3,573	65,428	5,723	2,141	3,307	70,414	5,145
Milieux ouverts	4,478	3,816	67,589	6,531	1,000	NA	70,000	NA
Prairies	1,198	1,746	63,729	4,671	1,163	1,948	65,049	4,424
Forêts résineuses	2,685	2,947	63,580	4,716	1,762	2,360	67,457	5,490
Moyenne globale	1,404	1,929	57,677	8,972	0,920	1,699	58,681	8,149
Moyenne hors ZC	1,173	1,590	55,293	8,428	0,803	1,475	55,865	7,469

a)

Trame	Fagne-Famenne				Lorraine			
	Alpha observé	Ecart-type	Alpha model	Ecart-type	Alpha observé	Ecart-type	Alpha model	Ecart-type
Milieux agricoles	2,015	2,276	68,162	3,272	NA	NA	NA	NA
Bocages	3,309	3,379	70,543	5,167	NA	NA	NA	NA
Cultures	1,500	1,737	67,646	2,369	0,769	1,180	65,641	2,367
Milieux aquatiques	4,519	5,699	78,099	6,719	3,895	4,350	76,047	5,623
Forêts feuillues	3,054	3,645	70,054	3,443	3,200	2,394	68,100	2,514
Forêts	2,972	3,401	70,632	3,882	2,786	2,259	67,286	2,525
Lisières	3,043	3,456	71,407	4,618	3,698	4,126	73,547	7,186
Milieux ouverts	6,074	4,038	73,778	8,768	4,100	4,706	78,129	4,773
Prairies	2,481	3,037	68,915	3,832	2,743	3,558	72,162	5,323
Forêts résineuses	3,624	4,295	72,028	5,386	3,579	2,589	68,632	4,946
Moyenne globale	2,061	2,873	65,526	9,322	1,471	2,180	63,165	9,417
Moyenne hors ZC	1,736	2,470	63,913	9,619	1,054	1,330	60,109	8,304

b)

Trame	Nord Sillon Sambre et Meuse			
	Alpha observé	Ecart-type	Alpha model	Ecart-type
Milieux agricoles	1,028	1,694	56,785	2,882
Bocages	0,862	1,270	57,972	3,496
Cultures	0,906	1,347	56,632	2,822
Milieux aquatiques	4,754	6,871	71,285	6,016
Forêts feuillues	3,357	5,624	67,269	7,064
Forêts	2,818	5,045	66,600	6,934
Lisières	4,202	6,309	70,246	6,670
Milieux ouverts	14,167	15,407	78,667	4,844
Prairies	1,261	1,962	61,042	5,150
Forêts résineuses	2,429	3,746	63,981	6,131
Moyenne globale	1,192	2,330	53,165	9,810
Moyenne hors ZC	1,003	1,750	50,166	8,986

c)

Pour l'intégralité des régions, l'indice alpha observé mesuré uniquement au sein du réseau des milieux de culture est à chaque fois inférieur à la moyenne globale de celui-ci. Il en est de même lorsqu'il est comparé avec la moyenne mesurée en dehors des zones cœurs. C'est la même chose pour la trame agricole en Fagne-Famenne et dans les autres régions sauf que sa valeur y est supérieure à la moyenne en dehors des zones cœurs. Elle n'est également pas représentée en Lorraine. Concernant les valeurs maximales, les trames associées sont assez variables en fonction de la région. Cependant, les trames des milieux ouverts, aquatiques et des lisières reviennent souvent dans le haut du classement. Les écarts-types pour toutes les régions confondues sont tous très élevés et dans la majorité des cas, sont supérieurs à la moyenne rendant les différences non significatives. En ce qui concerne le second indice, aucune valeur n'est inférieure à la moyenne globale. Cependant, la trame des milieux agricoles et ses sous-trames sont le plus fréquemment dans les dernières sauf au Condroz où les prairies enregistrent le troisième meilleur résultat. A contrario, les trames des milieux aquatiques et ouverts présentent les meilleures valeurs d'indice moyen. Les variances observées sont bien plus faibles que pour les deux autres indices et ceci dans chaque région.

## Résultats

Pour déterminer si les carrés kilométriques jugés comme populaires sont bien compris dans le réseau global, un test  $\chi^2$  est effectué sur un tableau de contingence (Tableau 22). Celui-ci s'est construit via la détermination d'un seuil différenciant les carrés kilométriques populaires des autres. Ce seuil s'élève à 21 observations suite à la classification de Jenks.

Tableau 22: Tableau de contingence montrant le nombre de zones cœurs populaires pour les ornithologues et permettant le test  $\chi^2$ .

	Populaire	Non populaire
Hors zone cœur	258	11816
Zone cœur	248	3910

Le test donne une p\_valeur inférieure à  $2,2 \cdot 10^{-16}$ , ce qui est hautement significatif. Le tableau 23 montre les valeurs qui été attendues si les deux variables (popularité et présence au sein d'un réseau) étaient bien indépendantes.

Tableau 23: Valeurs attendues si les deux variables testées étaient vraiment indépendantes.

	Populaire	Non populaire
Hors zone cœur	376,383	11697,617
Zone cœur	129,617	4028,383

Pour conclure cette analyse, le tableau 24 regroupe les résultats du calcul des corrélations réalisées pour les trois indices avec les occupations du sol. Comme lors de la sélection des espèces indicatrices, la normalité des distributions de l'indice alpha observé et du nombre d'observations par carré n'a pas été obtenue et aucune transformation ne semble pouvoir aider. Pour cela, une suppression systématique des données extrêmes en dehors de leur range a été réalisée. Par exemple, les carrés kilométriques ayant reçu plus d'une centaine d'observation ont été enlevés car ce nombre n'est clairement pas dû à l'occupation du sol et cela permet d'équilibrer la distribution.

Tableau 24: Valeur des corrélations entre l'indice alpha observé (a), alpha modélisé (b) et le nombre d'observations par carré kilométrique (c) pour les cinq biogéorégions de la Région Wallonne.

	Régions	Agricole	Bocage	Culture	Eau	Feuillue	Forêt	Lisière	Ouvert	Prairie	Résineux
a)	Fagne-Famenne	-0,033	0,089	-0,041	0,107	-0,059	-0,061	-0,005	0,326	-0,010	-0,023
	Lorraine	-0,056	0,079	-0,131	0,321	-0,074	-0,080	0,210	0,360	-0,007	-0,058
	Ardenne	-0,140	-0,047	-0,073	0,107	-0,016	0,009	0,132	0,286	-0,139	0,028
	Condroz	-0,120	-0,074	-0,063	0,192	0,088	0,094	0,125	0,158	-0,083	0,068
	NSSM	-0,108	0,103	-0,096	0,284	0,088	0,094	0,130	0,113	0,007	0,088
b)	Fagne-Famenne	0,273	0,198	-0,073	0,176	-0,364	-0,383	0,025	0,199	0,400	-0,180
	Lorraine	0,452	0,149	-0,009	0,390	-0,558	-0,571	0,075	0,458	0,520	-0,336
	Ardenne	0,384	0,140	0,143	0,203	-0,126	-0,452	-0,067	0,058	0,398	-0,465
	Condroz	-0,091	-0,025	-0,211	0,189	0,175	0,154	0,254	0,184	0,158	-0,011
	NSSM	-0,434	0,185	-0,559	0,419	0,451	0,444	0,511	0,256	0,417	0,157
c)	Fagne-Famenne	-0,034	0,087	-0,060	0,084	-0,045	-0,054	-0,021	0,354	0,004	-0,045
	Lorraine	-0,044	0,073	-0,111	0,236	-0,068	-0,078	0,150	0,365	0,001	-0,073
	Ardenne	-0,110	-0,044	-0,059	0,075	-0,010	0,008	0,106	0,225	-0,108	0,021
	Condroz	-0,056	-0,050	-0,031	0,162	0,016	0,018	0,068	0,128	-0,036	0,013
	NSSM	-0,034	0,057	-0,024	0,184	0,012	0,017	0,031	0,077	-0,012	0,044

Tout d'abord, l'analyse de l'indice alpha observé démontre que sur les trois premières régions, il possède les taux de corrélations les plus élevées allant de 0,36 en Ardenne à 0,326 dans les Fagnes concernant les milieux ouverts. Pour les deux restantes, les milieux aquatiques affichent le meilleur score avec seulement 0,192 au Condroz et 0,284 au nord du Sillon. Le second indice semble être bien plus influencé par l'occupation du sol.

En effet, la valeur maximale absolue enregistrée atteint -0,571 pour les forêts en Lorraine. Également, les cultures au nord du sillon possèdent une corrélation avec cet indice atteignant -0,559. La valeur maximale est atteinte en Lorraine concernant la trame des prairies avec 0,52. Le dernier indice montre un intervalle de valeurs assez similaire au premier. Ici, les milieux agricoles sont négativement corrélés avec et ceci, dans chaque région ainsi que pour la trame des cultures même si cette corrélation ne dépasse pas -0,11. Les prairies également, possèdent une corrélation négative mais pas en Fagne et en Lorraine. La corrélation maximale est atteinte dans cette dernière avec les milieux ouverts et vaut 0,365. Les forêts feuillues sont négativement corrélées avec les trois indices en Fagne-Famenne, Ardenne et Lorraine.

Les superficies des sites protégés par la LCN ou désignés comme SGIB recouverts par le réseau écologique global sont disponibles au sein du tableau 25. Il renseigne sur la proportion des surfaces sous statut de protection ou reconnues pour leur biodiversité remarquable et qui appartiennent également au réseau global. La proportion des surfaces sous protection qui appartient également à un réseau est aussi renseignée. Pour finir, la dernière colonne indique la proportion de surface commune entre le réseau existant (Figure 12) et le réseau potentiel global (Figure 15).

Tableau 25: Surface des sites mis sous statut de protection et couverts par le réseau global ainsi que la surface totale des sites sous statut et la proportion de surface de ces sites au sein du réseau global par rapport à leur surface totale.

Statut	Surface au sein du réseau (km <sup>2</sup> )	Surface totale (km <sup>2</sup> )	Proportion (%)	Proportion des réseaux sous protection (%)	Surface totale partagée Potentiel x Existant (km <sup>2</sup> )	
					Surface (km <sup>2</sup> )	Proportion (%)
LCN	67,861	143,2271	47,380	1,632		
SGIB	364,77	915,516	39,843	8,773		
Biodiversité remarquable	704	1658	42,461	16,931	971	23,353

Pour terminer, les cartes représentant les indices de diversités alpha ont été utilisées afin de caractériser chaque région biogéographique (Tableau 26). Ces mesures sont indépendantes aux réseaux écologiques précédemment délimités.

Tableau 26: Statistiques descriptives des valeurs des indices alpha suite à la modélisation et sur base des données empiriques.

Région	Indice Alpha modélisé				Indice Alpha empirique			
	Moyenne	Ecart-type	Maximum	Minimum	Moyenne	Ecart-type	Maximum	Minimum
Ardenne	57,677	8,972	89	30	1,404	1,929	16	0
Condroz	58,681	8,149	84	25	0,920	1,699	21	0
Fagne	65,526	9,322	91	37	2,061	2,873	32	0
Lorraine	63,165	9,417	91	42	1,471	2,180	27	0
Nord sillon								
Sambre et Meuse	53,165	9,810	88	23	1,192	2,330	36	0

Ces valeurs ne représentent que celles mesurées au sein d'un carré kilométrique afin d'avoir une idée de l'état de conservation de l'avifaune au sein de chaque région de Wallonie. Il semblerait que ça soit la Fagne-Famenne qui présente la meilleure moyenne de présence modélisée par carré suivie de très près par la Lorraine et ensuite par le Condroz. Cependant, en vue de leur variance relativement élevée, aucune différence significative ne peut être prouvée entre ces régions. Ceci est encore plus vrai concernant l'indice alpha non modélisé où les écarts-types sont supérieurs à la moyenne. Le classement reste le même, mis à part que le Condroz passe cette fois en dernier.

## IV. Discussion

### 4.1. *Acquisition et traitement des données*

Parmi toutes les espèces classées comme rares, seules les deux espèces du genre *Milvus* n'ont pas été modélisées malgré un nombre suffisant de points de présence. La raison est que les points de présence de ces deux espèces ne proviennent pas d'observations courantes, mais sont les résultats d'un inventaire réalisé entre l'année 2015 et 2016 à une échelle de 40km<sup>2</sup>. Cette différence d'échelle est l'élément majeur empêchant la réalisation d'une modélisation correcte. En effet, approximer les points de présences de ces deux espèces à l'ensemble des points centraux des carrés kilométriques constituant les surfaces où l'espèce a été recensée apporterait un biais trop important. Concernant le Tarier des prés, il est important de rester prudent par rapport au modèle final de celui-ci. En effet, les points de présences sont déterminés via les carrés kilométriques se trouvant complètement à l'intérieur des sites Natura 2000 où l'espèce a été inventoriée. Ceci a pu amener à une surestimation de la niche écologique potentielle de l'espèce due à un taux de fausses présences supérieur à la normale. Cependant, les sites possèdent pour la plupart une superficie inférieure à 1km<sup>2</sup> limitant le problème d'obtenir trop de fausses présences contrairement aux Milans. Cet effet négatif pourrait engendrer une surestimation des aires de distribution de l'espèce modélisée.

### 4.2. *Révision des trames*

La révision de trames avant modélisation a eu deux effets importants. Le premier est l'ajout de la trame résineuse et le second, la modification et la clarification des trames secondaires des lisières et bocages ainsi que des forêts. L'objectif du réseau écologique est de protéger les zones principales de nidifications de l'avifaune en Région Wallonne. De ce fait, une trame des milieux artificiels n'a pas vraiment de raisons d'être, car les sites fortement urbanisés seraient à éviter. Cependant, cette trame thématique n'est pas sans importance, car de nombreuses espèces sont tout de même affiliées à ces milieux. Il serait plus avisé de réaliser une étude plus précise dans ces zones afin de sélectionner plus précisément les sites d'intérêt à l'intérieur des villes. Concernant les milieux rocheux, peu d'espèces semblaient liées à cette trame à la base. Ensuite, la quasi-totalité d'entre elles est également affiliée aux milieux aquatiques et particulièrement aux cours d'eau dus sans doute, aux affleurements rocheux qu'ils créent. La couche écotope utilisée à cette résolution pour la représenter ne semble également pas idéale. En effet, il faudrait une résolution bien plus précise afin que les éléments rocheux deviennent perceptibles et analysables.

Les autres trames sont quant à elles relativement simples à matérialiser via les variables écotopes décrivant l'occupation du sol liée à une trame. Ceci a été plus complexe pour la trame « Bocages et lisières » où une séparation est nécessaire pour des analyses ultérieures. Il a été décidé d'assimiler les lisières à la variable « Copen ». Cette variable représente les milieux perturbés qui englobent par conséquent les zones recouvertes de buissons et d'herbacés formés par de jeunes arbres ou des adventices forestiers. Les surfaces des lisières sont toujours assez complexes à matérialiser, spécialement à une échelle relativement grossière du fait que selon la définition, celles-ci possèdent une épaisseur fortement variable. En effet, une lisière représente la surface au sein d'une tache forestière qui est influencée par la matrice paysagère. Selon les variables, les espèces ou encore le massif forestier pris en compte, l'épaisseur peut varier énormément (Camargo et al., 1995; Dupont et al., 2011).

Une assimilation de cette trame à la variable Copen semble être la meilleure solution, car les milieux perturbés comme des trouées, de jeunes plantations ou encore des mises à blanc partagent de nombreuses caractéristiques écologiques avec les lisières. De ce fait, une espèce indicatrice d'un de ces habitats serait également indicatrice de l'autre. Pour finir, l'analyse de la trame bocagère ne se base pas sur une variable écotope comme précédemment, mais sur une carte construite sur base de photointerprétation représentant la délimitation des bocages. C'est pourquoi le niveau de précision n'est pas aussi bon que celui des analyses basées sur les écotopes de par sa nature plus arbitraire.

### 4.3. *Modélisation*

Les variables environnementales utilisées lors de la modélisation des distributions de chaque espèce ont été choisies de manière à réduire leur corrélation au maximum. C'est pourquoi plus de la moitié des variables écotopes issues de Wordclim ont été retirées. En effet, toutes celles relatives aux climats sont sensiblement corrélées entre elles. Également, le climat en Belgique est fortement influencé par l'altitude ce qui explique la suppression de la variable « élévation ». Cependant, certaines corrélations sont cryptiques et indirectes sans pour autant que leur suppression ne devienne facultative. Un bon exemple à cela est la variable artificiel et Light qui décrivent le même phénomène qui est le degré d'anthropisation d'un milieu, mais la méthode de sélection les a prises toutes les deux dans le jeu de variables final. La première est dans ce cas, maintenue, car elle est la plus à même à décrire ce phénomène. Les variables « Azimut » et « SunSpring » ne sont également pas identifiées comme corrélées l'une à l'autre car elles ne sont pas graduées de la même façon. En effet, l'azimut est mesuré en degré alors que Sunspring est en Watt par m<sup>2</sup>.

Les points de présence de chaque espèce sont résumés au point central du carré kilométrique dans lequel ils se trouvent. Cela pourrait induire une certaine perte d'informations à première vue, car de nombreux points de présence sont enlevés, mais étant donné la nature des données, il est bien plus avisé de procéder ainsi. En effet, ne considérer qu'un seul point par carré kilométrique diminue nettement le poids des zones ayant été suréchantillonnées. Ceci réduit par conséquent le biais d'échantillonnage qui est relativement important avec des données « civiles ». Également, l'autocorrélation spatiale pouvant exister entre les points de présence ainsi que l'apparition de pseudorépliques se limite fortement. Ce phénomène apparaît lorsqu'un individu a été observé deux fois causant l'apparition d'un deuxième point d'observation pour le même individu (Guisan et al., 2017). L'autocorrélation entre les points de validation et de calibration diminue également (Veloz, 2009) ce qui améliore la pertinence des indices d'évaluation se basant sur eux.

#### 4.3.1. Méthode proposée

La deuxième étape clé de la modélisation est de réduire au maximum le biais d'échantillonnage. Celui-ci regroupe pour rappel tous les phénomènes ayant un impact négatif sur la qualité des échantillons pris en compte qui sont ici, les points de présences. Au plus ce biais est important et au moins les observations ainsi que les analyses de celles-ci seront fiables car la véracité des données sera également plus faible. Dans ce travail, des méthodes d'amortissement et de réduction du biais ont été testées et se concentrent sur la sélection des pseudo-absences afin de minimiser le taux de fausses absences. Des points d'absence sûrs sont difficiles à obtenir sans la réalisation d'inventaires exhaustifs. C'est pourquoi, en vue du grand nombre d'observations disponibles, il a été décidé d'utiliser le nombre d'observations par carré kilométrique comme proxy de l'effort d'échantillonnage.

Il est assez facile d'admettre qu'au plus une zone a été fréquentée et qu'une espèce y est absente, au plus la certitude de celle-ci est élevée. Cependant, cela se base sur une hypothèse forte qui stipule que chaque espèce serait recensée si elle est présente lors de l'observation d'une autre espèce d'oiseau. Également, il y a une telle variabilité au sein de la multitude d'observateurs qu'il est impossible de prédire parfaitement leur attitude lors de leurs observations. De plus, cela peut réduire le nombre de points d'absences pour les espèces assez répandues. Une seconde méthode semblable a également été testée qui consiste à utiliser la richesse spécifique d'un carré kilométrique comme proxy. La même hypothèse forte est également à la base de ce raisonnement impliquant les mêmes points négatifs. Cependant, une certaine prise en compte du comportement des observateurs est toute de même faite. En effet, au lieu de se baser sur le nombre d'observateurs par carré sans s'informer sur leurs observations, ici, elles sont directement prises en compte. Malgré tout, il aurait été nettement préférable d'utiliser le nombre d'observations réalisées durant la même journée ou par le même observateur. Ces alternatives possèdent une bien meilleure pertinence, mais ne sont pas possibles à mettre en place en raison d'informations manquantes pour de nombreuses espèces et d'un intervalle de temps d'observation variable. Malgré tout, ces deux techniques possèdent des avantages ne pouvant être ignorés.

Un autre bon moyen de gérer ce biais ou du moins de le rendre moins dommageable est d'égaliser le nombre de présences et d'absences. En effet, s'ils s'égalisent, les biais peuvent s'autocompenser lors de l'estimation de nouveaux points de présences et d'absence s'ils sont à peu près équivalents. Malheureusement, certaines espèces telles que le Petit gravelot ou le Martin pêcheur possèdent juste assez de points de présence pour construire un modèle assez robuste. Si le nombre d'absences diminue de trop, cela risque de créer des modèles bien trop généreux comme il a pu être observé durant des tests préparatoires et comme Vanderwal et al.,(2008) le stipulent.

#### 4.3.2. Indice et méthode de sélection

La majorité de la variabilité entre les espèces est englobée grâce à une sélection d'espèces assez différentes l'une de l'autre.

En absence de seuil séparant les données positives et négatives, le moyen le plus commun afin d'évaluer un modèle est bien l'AUC de la ROC (Chicco et al., 2020). Cependant, il ne s'agit pas de l'indice le plus fiable. En effet, il est assez sensible aux échantillons dont la prévalence est différente de 0,5. Ceci est d'ailleurs le cas pour les trois méthodes. C'est donc un élément impératif à prendre en compte car certains indices deviennent inutilisables ou leur interprétation est modifiée. C'est le cas de l'AUC, mais surtout de l'UA qui dans sa façon d'être calculé induit un déséquilibre.

L'un de ses avantages par contre, est qu'il permet de sélectionner le seuil manquant permettant la mesure d'autres indices basés sur une matrice de confusion. Cependant, le problème de la prévalence reste, c'est pourquoi à la place d'une mesure de la précision globale du modèle, c'est l'indice de corrélation de Matthew qui est calculé. L'indice classique est énormément biaisé en faveur de la catégorie la plus répandue qui dans ce cas-ci est les absences. Le second est censé être indépendant à ce phénomène. Ensuite, deux autres indices de précisions sont utilisés, mais cette fois pour évaluer le taux d'omission via le PA et celui de commission via l'UA. Malheureusement, ce dernier s'avère très sensible au déséquilibre entre le nombre de présences et d'absences. En effet, au plus il y aura d'absences et au plus les probabilités d'erreur de commission seront élevées.

Il est facilement observable au sein du tableau 8 que les valeurs de l'UA de la deuxième et troisième méthode sont nettement supérieures à celle avec un background de 10 000 points pris au hasard. Cette différence de valeur ne s'explique pas par une qualité supérieure, mais par la différence de nombre de points d'absence. Ceci empêche l'utilisation de cet indice lors de la sélection de la meilleure méthode car la comparaison n'est pas pertinente. Il s'agit d'ailleurs d'un défaut récurrent au sein des indices de précision, car ils sont fortement biaisés par la prévalence (Powers, 2011). Malgré tout, la différence dans le nombre de points d'absences est plus faible entre la deuxième et troisième méthode permettant tout de même une comparaison. À deux reprises, pour le Pic épeichette et le Hibou moyen-duc, la deuxième méthode est meilleure la rendant visiblement plus apte à identifier les points d'absence.

Concernant l'erreur d'omission, aucune différence significative n'apparaît entre les espèces. À un seul moment, la première méthode est clairement inférieure aux deux autres concernant l'Autour des palombes. Pour le Faucon hobereau, il s'agit de la deuxième méthode qui est nettement inférieure aux autres. Ainsi, la seule conclusion pouvant être amenée est que la troisième offrirait une meilleure prédiction des points de présence sans que cela soit suffisant pour la sélectionner comme méthode finale.

Dans l'entièreté des cas, à l'exception du Faucon hobereau, la première méthode obtient des résultats significativement inférieurs aux deux autres pour le coefficient de Matthew. Une démarcation entre celles-ci n'apparaît que pour le Hibou moyen-duc et le Pic épeichette en faveur de celle basée sur la richesse spécifique. Comme il s'agit de la représentation de la précision globale, cela indique que pour le Faucon hobereau, la deuxième méthode se rattrape grâce à une bonne prédiction des points d'absence alors que la troisième méthode n'y parvient pas grâce à sa meilleure prédiction des points de présences.

Objectivement, la deuxième méthode paraît être la plus adéquate et la première serait à éviter. Cependant, les méthodes 1 et 2 basent leur background sur le principe appelé « Target group » faisant apparaître une autocorrélation entre les points d'absence. Même si l'utilisation d'un grand nombre d'espèce limite celle-ci, elle existe toujours. Dû à cela, les données de calibration et de validation sont moins indépendantes ce qui augmente les résultats des indices relatifs aux absences sans que cela soit écologiquement justifié. Ce phénomène pourrait être à l'origine de l'infériorité générale de la précision de la première méthode. Cependant, l'AUC est censé au même titre que les autres indices, être gonflé dû à cette autocorrélation des points d'absence devant logiquement amener un meilleur résultat pour les deux dernières méthodes (Iturbide et al., 2015). Or, la deuxième est significativement supérieure à la première qu'une seule fois pour le Hibou moyen-duc où elle avait déjà eu le meilleur coefficient de Matthew. La première quant à elle, est nettement supérieure aux deux autres à deux reprises. À ce niveau-ci et en prenant en compte tous les éléments discutés plus haut, le choix ne semble plus si évident. En effet, la supériorité apparente des deux méthodes aux absences améliorées au niveau des indices de précision est majoritairement expliquée par cette dépendance entre les points de calibration et de validation. Ensuite, les AUC de ces deux méthodes devraient être également supérieures pour les mêmes raisons, mais cela n'est significativement pas le cas. Pour assurer une sélection valable de la bonne méthode, il est plus avisé d'également regarder la représentation cartographique des différents modèles et de les comparer.

Il a déjà été remarqué auparavant (Iturbide et al., 2015) que la technique du « Target group » cause l'estimation d'une trop grande aire de distribution spécialement dans les zones non ou faiblement échantillonnées. Également, le phénomène inverse apparaît lorsque toutes les données disponibles sont utilisées surtout lorsque la modélisation se fait sur une grande aire d'étude où les variabilités environnementales sont plus nombreuses (Broennimann et al., 2008).

Cela peut être observé dans ce cas d'étude en comparant les différentes niches écologiques estimées par chaque méthode pour chacune des dix espèces. En effet, au plus le nombre d'absences prises en compte diminue et au plus les modèles sont généreux. C'est pourquoi la méthode trois possédant légèrement plus de points d'absence semble offrir des résultats plus réalistes que la deuxième tout en gardant une qualité inférieure à la première. Pour confirmer et finaliser cette sélection, un avis d'expert est nécessaire et celui-ci penche en faveur de la première méthode qui semble offrir des niches écologiques potentielles bien plus réalistes.

Avec ce choix, quelques conclusions pourraient être proposées par rapport aux résultats obtenus. Il semblerait que malgré son utilisation pour compenser les défauts de l'AUC, l'indice de corrélation de Matthew soit encore plus sensible à l'autocorrélation des données. En effet, sans que la première méthode ne montre un taux d'erreurs d'omission significativement plus faible que les deux autres méthodes en analysant leur PA, celle-ci possède une précision globale nettement inférieure. Cela amène par déduction à penser qu'elle possède un taux d'erreurs de commission nettement plus élevé que les autres. Or, elle est la seule à offrir des aires de distribution réaliste. Cette surestimation devrait mener naturellement à une hausse du taux d'erreur de commission qui n'est visiblement pas identifiée par l'indice de Matthew principalement à cause de cette autocorrélation. Le fait que l'AUC de ces deux méthodes ne montre pas de supériorité apparente n'assure pas qu'il n'a pas été influencé par l'effet négatif des Target groups. L'hypothèse la plus probable serait que le gonflement qui aurait dû être provoqué selon la littérature (Veloz, 2009) s'est tout de même produit et a servi à combler l'écart de qualité apparente entre ces deux méthodes et la première. Ainsi, l'effet du nombre de données assez réduit aurait diminué la vraie qualité du modèle, car certaines variables d'exclusion ne sont peut-être plus prises en compte. L'autocorrélation existante aurait quant à elle, augmenté la qualité apparente des modèles.

Ces deux techniques semblent tout de même prometteuses et pourraient offrir de bons résultats si le jeu de données de l'entière des espèces était plus grand ou s'il y avait un plus grand nombre de points d'absences sélectionnables. Également, elles pourraient s'avérer très utiles afin d'utiliser un algorithme nécessitant de vraies absences comme GAM. C'est pourquoi la pertinence de cette hypothèse a été testée et les modèles obtenus comparés avec la méthode par défaut de MaxEnt. Étant donné tous les problèmes de comparaison rencontrés précédemment qui seraient à nouveau présents, seul l'AUC est utilisé pour cette évaluation.

À première vue, les résultats sont assez bons ce qui laisserait supposer que les deux techniques d'amélioration des absences tiennent la route. Pour pas mal d'espèces, au moins l'une des deux méthodes obtient un meilleur AUC que ceux obtenus via MaxEnt. Cependant, cette qualité supérieure est loin d'être systématique et les valeurs sont fort variables d'une espèce à l'autre. Elles proposent aussi des modèles trop généreux. Également, cet algorithme est plus sensible aux échantillons de petite taille offrant une qualité de modèle beaucoup trop variable en fonction de l'espèce modélisée. C'est pour ces différentes raisons que le reste des modélisations se sont réalisées avec MaxEnt.



### 4.3.3. Modélisation finale

Même si précédemment, les indices choisis pour l'évaluation des modèles ont montré quelques défauts, ces derniers étaient spécifiques à la situation. À présent, ils servent à choisir le set de variables le plus adéquat à estimer la distribution d'une espèce. Les nombres de points d'absence et de présence restent constants tout comme leur localisation tout en ayant une indépendance plus forte entre les points de calibration et de validation. Une fois de plus, une attention plus particulière à la réduction de l'erreur de commission a été faite via l'UA.

La mesure de l'inverse de l'erreur de commission offre des résultats nuancés mais en général, assez bas pour la totalité des espèces. Ceci s'explique par la présence d'un biais dans la méthode de calcul de cette erreur. Elle ne donne donc pas uniquement le taux de fausses présences. En effet, il s'agit plus de la mesure d'un indice de commission et serait composé du taux et d'un autre terme. Ce dernier apparaît à cause de la prévalence nettement inférieure à 0,5 et plus précisément du peu de points de présence de validation disponibles. Le modèle pourrait estimer la présence d'une espèce dans une zone où aucune présence n'y a été observée ne permettant pas de la confirmer ce qui élève le nombre de fausses présences. Les niches écologiques modélisées de cette manière représentent idéalement la niche fondamentale de l'espèce qui se réduit intrinsèquement à la sa niche potentielle. De cette manière, certaines zones identifiées comme adéquates pour l'espèce peuvent ne pas l'abriter pour des raisons physiques ou historiques et non biologiques (Anderson et al., 2003). En résumé, trois cas de figure peuvent arriver, soit il s'agit bien d'une erreur de commission où l'espèce n'est pas présente et ne pourrait écologiquement pas y être, impliquant une faiblesse au sein de modèle. Soit, l'espèce y est bien présente, mais le manque de données empêche la validation du point. Le dernier cas est que l'espèce n'y est effectivement pas présente, mais pour des raisons dépassant les variables utilisées n'impliquant pas directement une faiblesse du modèle. Malgré tout, il est tout de même bien de maximiser l'indice afin de minimiser une éventuelle surestimation de la distribution de l'espèce.

Le fait de ne pas pouvoir comparer cet indice entre espèces empêche de déterminer un certain seuil qui distinguerait les modèles de mauvaise qualité. Malgré tout, les différences de prévalence sont beaucoup moins prononcées qu'à l'étape précédente, c'est pourquoi une certaine comparaison peut tout de même être faite. Il faut garder à l'esprit qu'un UA inférieur à celui d'un autre modèle ne signifie pas forcément un plus haut taux d'erreur de commissions. De plus, l'indice de précision globale devrait pouvoir être comparé entre les modèles d'espèces différentes donnant une meilleure capacité d'évaluation de ces derniers.

Toutes les espèces de la liste rouge (Tableau 10) ont été écartées lors de la sélection des espèces indicatrices et de la mesure de l'indice alpha pondéré en raison de la qualité hasardeuse de leur modèle. Cela a été fait ainsi surtout par prudence afin d'éviter une sélection de zones cœurs trop grossière et généreuse. Cela ne veut pas dire que tous les modèles cités dans ce tableau sont à jeter et que les autres sont parfaits. Comme précédemment, la priorité du jugement a été mise sur le coefficient de Matthew et l'UA afin de réduire au maximum les éventuelles surestimations des zones d'habitat potentiel. Cependant, la majorité des espèces du tableau ne compte pas plus d'une trentaine de points de présence pour valider le modèle causant une surestimation de l'erreur de commission. C'est pourquoi parmi eux, une seconde séparation a été réalisée différenciant les modèles indignes de confiance en tout point et ceux ne possédant juste un mauvais UA. Parmi ceux-là, il y a par exemple le Martin pêcheur dont la plupart de ses indices sont acceptables. Également, certaines espèces dont leur modèle a été désigné comme bon ne sont tout de même pas entièrement dignes de confiance. Certains possèdent des indices ayant des valeurs relativement faibles, mais pas assez pour être classés sur liste rouge.

Pour revenir aux quatre espèces ayant un modèle de très basse qualité, la raison principale à cela est qu'ils possèdent un nombre de points de présence relativement grand allant de 479 pour la Bondrée et 627 pour le Hibou. Or, cela devrait augmenter un minimum la valeur des indices sans être le cas. Les deux espèces de faucons, quant à eux, possèdent moins d'une centaine de points de présence justifiant un peu plus cette qualité médiocre. Cependant, la valeur de leur PA est juste au-dessus de 0,5 indiquant que leur modèle parvient à identifier les points de présence légèrement mieux que si cela avait été fait au hasard. Autrement dit, le taux de présences non identifiées (d'omission) est proche de un sur deux ce qui est très mauvais même lorsque l'erreur d'omission n'est pas la plus problématique. A contrario, un total de cinq espèces offre de très bons modèles avec une valeur élevée pour l'entière des indices mesurés. Pour la grande majorité, elles sont des espèces assez spécialisées à un habitat ce qui facilite leur modélisation sans nécessiter un grand nombre de données d'entrée. D'ailleurs, les espèces dont l'AUC de leur modèle est relativement grand sont plus susceptibles d'être sélectionnées comme indicatrices (Lobo et al.2008).

Pour finir, l'AICc a très peu été pris en compte lors de l'évaluation finale des modèles et de leur comparaison. Dans la plupart des cas, il est important, voir même indispensable de réduire la complexité d'un modèle afin d'éviter un surajustement des données et obtenir des résultats trop spécifiques au set de variables locales. Cependant, cela est surtout important lorsque le modèle en question sera projeté sur d'autres aires d'études ce qui n'est pas le cas ici. Au contraire, comme un modèle est censé de mieux en mieux ajuster une distribution au plus il devient complexe, lorsqu'il décrit la distribution d'une espèce au sein d'une aire bien déterminée, il gagne en qualité. C'est pourquoi de nombreux modèles ont été construits avec l'entière des variables écotopes non corrélées.

## 4.4. *Construction des réseaux écologiques*

### 4.4.1. Sélection des espèces indicatrices

Généralement, lors de l'identification d'espèces indicatrices, l'indice Indval est utilisé (Dufrêne et al., 1997). La résolution et les variables environnementales choisies dans le cadre de cette étude ne permettent cependant pas l'utilisation de celui-ci. En effet, comme les carrés kilométriques sont difficiles à classer par trame écologique sans créer de biais, il faut directement utiliser les données quantitatives offertes par les variables écotopes. Grâce à celles-ci, il suffit de trouver un lien entre les surfaces d'occupation du sol relatives à une trame et la distribution de chaque espèce.

Grâce à la séparation des analyses en fonction des territoires biogéographiques et de la suppression de certaines espèces en conséquence, les représentations graphiques des RDA suivent toutes la même logique. En effet, les axes des variables s'arrangent globalement de la même manière et de façon logique pour les graphes où toutes les trames sont prises en compte. Ainsi, les forêts feuillues et résineuses se trouvent toujours l'une près de l'autre entourées par la trame bocagère et celles des lisières. Cette dernière est suivie par la trame des milieux ouverts elle-même succédée par la variable « Artificielle » qui fait place aux milieux agricoles. Le graphe se ferme via la jointure entre celle-ci et celle des bocages. Il y a tout de même certaines modifications de ce schéma global qui peuvent apparaître en fonction de la région considérée. Par exemple, la trame des milieux aquatiques se positionne proche des forêts feuillues en Ardenne car il s'agit d'une région principalement boisée et se situera proche des milieux ouverts et agricoles au nord du sillon Sambre et Meuse (Figure 7).

Cependant, aucune véritable conclusion ne peut être tirée de ces analyses globales, les variables d'occupation du sol doivent être prises une par une.

Comme il a été expliqué plus en détail précédemment, le « goodness of fit » est la donnée la plus intéressante dans ce cadre-ci. Par ailleurs, des seuils de sélection différents entre les espèces rares et communes ont été fixés par rapport à cette mesure. Cela a été décidé après la réalisation de l'entièreté des analyses en vue des résultats obtenus s'avérant plus faibles pour les espèces rares dues probablement à la différence des données utilisées. D'un côté, il y a les probabilités de Maxent où les seuils entre une absence et une présence sont nettement inférieurs à 0,5 ce qui a pu causer une sous-estimation des relations entre espèces et variables. Par ailleurs, la somme des goodness of fit d'une espèce par région peut être supérieure à 100% démontrant la nécessité d'une analyse commune des résultats de chaque trame.

La seconde sélection plus contraignante a permis d'assurer la pertinence des résultats des RDA sans pour autant baser la sélection uniquement sur l'avis d'experts. Il s'agissait plus de faire le tri entre les espèces indicatrices pour ne pas avoir une dizaine d'espèces par trame et de ne prendre que les plus pertinentes. De plus, en vue des espèces choisies, cette méthode semble assez efficace et précise.

En suivant ce procédé, la séparation des milieux naturels ouverts en milieux secs et humides ne donne aucune espèce représentative de cette dernière et ceci, dans à peu près toutes les régions. La variable « Wopen » qui fut utilisée pour représenter cette trame regroupe les terrains inondés à dominance herbacée tels que les tourbières. Il s'agit généralement d'habitat d'une importance considérable en matière de biodiversité. Cependant, à cette résolution, ils risquent fort de passer inaperçus. La relation entre l'abondance ou la probabilité de présence d'une espèce et cette densité surfacique pourrait donc être masquée.

La trame des milieux rocheux n'a également pas été maintenue pour les mêmes raisons, car la variable utilisée pour la représenter n'offre pas une gamme de valeurs différentes suffisamment élevée. Ainsi, les relations entre la présence d'une espèce et les valeurs que prend cette variable sont assez compliquées à obtenir et ne donneraient pas de conclusions pertinentes.

La séparation des analyses en fonction des territoires biogéographiques et ainsi du gradient d'altitude semble avoir porté ses fruits en voyant que les espèces indicatrices d'une trame ne sont pas constantes en fonction des régions. Par exemple, la Chouette de Tengmalm est connue pour se concentrer principalement dans les forêts ardennaises et donc, serait une bonne espèce indicatrice dans cette région ce qui est d'ailleurs le cas. Par ailleurs, si l'analyse avait été faite de façon globale sur l'entièreté du territoire, elle n'aurait pas été sélectionnée.

En analysant plus précisément le lien entre l'AUC et le caractère indicateur d'une espèce, il s'avère que plus de 85% des espèces indicatrices possèdent un AUC supérieur à 0,8. Ainsi sur les 21 espèces rares indicatrices, seules trois d'entre elles ont un AUC moyen ou mauvais (*Tyto alba*). Un AUC élevé est déjà un bon indicateur de la spécialisation d'une espèce envers un milieu spécifique.

#### 4.4.2. Délimitation des zones cœurs

##### 4.4.2.1. Réseau existant

Le réseau existant a pour objectif de représenter les sites étant déjà reconnus comme enjeux majeurs pour la conservation de la nature. Il regroupe donc les sites Natura2000 importants, les SGIB et les ceux protégés par la LCN. Initialement, il aurait été mieux de prendre en compte uniquement les unités de gestion prioritaires de Natura 2000 sauf qu'ils sont majoritairement déjà désignés comme SGIB. À la place, il a été décidé de mettre plus d'importance aux sites concernés par la LCN et les SGIB via une pondération les favorisant. Ces sites ne sont par contre, pas délimités à l'échelle de 1km<sup>2</sup> ce qui demande la détermination de seuils pour augmenter cette résolution. Cette technique a été choisie car elle est simple à mettre en place, spécialement sur l'étendue de toute la Région wallonne et efficace.

Son défaut majeur est qu'elle pourrait sembler relativement arbitraire car le nombre de classes choisies lors de la classification de Jenks n'a pas vraiment de fondements scientifiques solides. Cependant, la logique de création de deux groupes ayant la plus grande variance entre eux reste assez bonne tout comme les résultats obtenus. Également, il s'agit plus d'une représentation cartographique (Figure 12) et qu'il serait bien plus prudent d'effectuer des analyses directement sur les surfaces des sites sous statuts.

Suivant le même principe, il est nécessaire d'ajouter les carrés kilométriques ayant une diversité d'oiseaux remarquables et déjà reconnus via les observations des citoyens. Pour cela, une simple utilisation de l'indice alpha observé est faite. Contrairement aux précédentes cartes, il est important d'étudier la diversité des carrés kilométriques séparément par régions biogéographiques. En effet, celle-ci dépend fortement de l'état écologique de la région en question. De cette manière, un seuil plus élevé apparaît en Fagne-Famenne au contraire du Condroz. Via ces seuils spécifiques, les différents carrés kilométriques de chaque région ayant une diversité remarquable peuvent être discernés des autres. Malheureusement, cette évaluation n'a pu se faire uniquement sur les observations des espèces dites rares. Malgré tout, le contraire aurait été plus dommageable, car la présence d'une espèce rare a plus de poids dans la caractérisation de la biodiversité d'un milieu.

#### 4.4.2.2. Réseau potentiel

Au point précédent, les données de présence modélisées ne sont pas prises en compte, car il s'agit d'estimation plutôt que de véritables observations. Cependant, ces données ont l'avantage de prédire les zones de présence dans des milieux non ou mal inventoriés (cfr Introduction). De cette manière, elles permettent de déterminer les sites importants qui ne sont pas encore connus. L'objectif principal du réseau potentiel est bien de mettre en avant les zones importantes pour la nidification des oiseaux de Wallonie indépendamment de la législation déjà mise en place et des observations directes.

L'indice alpha mesuré sur les présences modélisées représente le meilleur indice de diversité permettant sa caractérisation et la détermination des sites d'intérêt. Au contraire de l'indice alpha précédent, celui-ci comptabilise la présence de l'entièreté des espèces nicheuses en Wallonie. Cependant, celles n'ayant pas pu être modélisées (Annexe 3) en sont tout de même absentes ainsi que les espèces ayant un modèle de basses qualités ce qui amène un léger biais. Les seuils de différenciation concernant cet indice ne donnent pas exactement le même ordre croissant que le précédent. En effet, un plus grand nombre d'espèces est attendu dans les carrés kilométriques désignés en Ardenne qu'au nord du Sillon. Il semblerait donc que les données modélisées ne soient déjà pas en accord avec celles des observations directes causant une différence dans le classement des régions biogéographique en matière de richesse en oiseau. Cependant, la Fagne-Famenne maintient sa supériorité considérable dans les deux cas.

Une fois le nombre d'espèces modélisées dépassé, les différents carrés kilométriques ainsi sélectionnés doivent être affiliés à une ou plusieurs trames écologiques selon les espèces qui s'y trouvent. En effet, une échelle aussi grande ne peut limiter l'appartenance d'un carré à une seule trame. Initialement, il était prévu d'uniquement compter sur la présence d'au moins une espèce indicatrice pour déterminer les zones cœurs de biodiversité. Cependant, baser la classification écologique d'une telle surface rien que sur cela amène une surestimation trop forte. La solution alternative choisie a été de réutiliser les résultats bruts des RDA afin de générer des cartes spécifiques à chaque trame. Celles-ci représentent la diversité spécifique de chaque carré kilométrique par rapport à la trame en question, c'est-à-dire qu'au plus les valeurs seront élevées et au plus des espèces affiliées à celle-ci y seraient présentes.

Cependant, une grande valeur d'indice alpha pondéré peut signifier plusieurs cas. Le plus idéal serait le cas d'une présence importante d'espèces affiliées à la trame. Cela pourrait également signifier la présence de nombreuses espèces neutres. Cependant, comme les RDA ont visiblement bien fonctionné et que la contrainte de la présence d'au moins une espèce indicatrice est maintenue, il semblerait que le premier cas soit le plus fréquent. Cette manipulation a permis d'affiner les réseaux par rapport à la trame et d'en améliorer sa précision. Il semble plus avisé de procéder ainsi que de directement classer les carrés kilométriques en fonction de leur occupation du sol. En effet, cette méthode pourrait enlever systématiquement des zones ayant une faible proportion d'occupation par cette trame, mais qui abriteraient tout de même de nombreuses espèces inféodées à celle-ci. En utilisant l'indice pondéré, ce phénomène est moins probable, car la présence d'une espèce indicatrice offre un poids bien supérieur que la surface prise seule. Ici, c'est la distribution des espèces ainsi que leur niveau d'affiliation à la trame qui permet de les cartographier. Ceci a permis par exemple pour la trame des forêts feuillues de prendre en compte un nombre de carrés kilométriques supérieur lors de la sélection finale. Cependant, seuls les résultats des espèces rares ont été utilisés, car les cartes obtenues après pondération ne reflétaient pas assez précisément les couches d'occupation du sol de la trame. Il était donc bien plus prudent de ne se baser uniquement sur les espèces rares.

La majorité des réseaux écologiques ainsi construits sont tout de même de superficie assez importante, mais en travaillant séparément par rapport aux régions, les trames sont toutes plus ou moins bien représentées dans chacune d'entre elles. Ainsi, même si le nord du Sillon et le Condroz sont dominés par les milieux agricoles, un bon nombre de zones forestières sont tout de même identifiées comme zone cœur. D'ailleurs, la plus grande rareté d'une certaine occupation du sol dans une région donne encore plus d'importance à la protection de celle-ci.

Une autre alternative relativement simple aurait été de poser un seuil sur le nombre d'espèces indicatrices nécessaires à la sélection finale d'une zone cœur. De cette manière, la représentativité de la trame concernée au sein du carré kilométrique serait nettement renforcée. Également, afin de limiter les présences fortement rependues des espèces communes, une pondération sur celle-ci de la forme « *présence pondérée* =  $\frac{1}{\ln(\sum \text{présence})}$  » serait conseillée. Cette méthode a été testée pour la trame des prairies en Ardenne où sa distribution est assez bien nuancée (Figure 18).

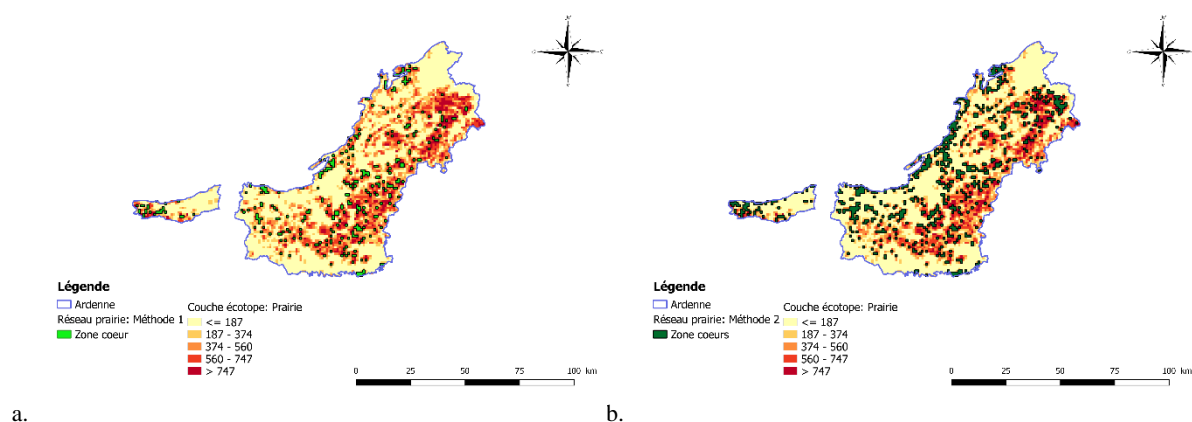


Figure 18: Cartographie des zones cœurs de la trame des prairies en Ardenne via la première méthode (a.) et la méthode alternative (b.).

Lorsque les pixels résultants de cette classification sont classés de la même manière que pour l'alpha pondéré, un seuil de 0,524 est déterminé. Ainsi, lorsqu'il est appliqué aux nombres d'espèces pondérés, un nombre supérieur de zones cœurs à la méthode initiale est déterminé. Or, l'objectif escompté aurait été de limiter ce nombre. Également, les deux techniques permettent une sélection des carrés kilométriques ayant le plus de surfaces de prairies montrant une bonne précision pour les deux. Cependant, l'utilisation de plusieurs espèces indicatrices pose problème pour les trames qui dans certaines régions ne disposent que d'une seule. Le cas le plus dommageable semble être pour les milieux bocagers qui ne possèdent qu'une espèce indicatrice au nord du Sillon et dans le Condroz. Cette espèce est d'ailleurs fortement répandue ce qui augmente énormément le nombre de zones cœurs sélectionnées sans qu'il n'y ait de bocages justifiant cette sélection. Ce phénomène apparaît car les résultats des RDA ne montrent pas de fortes affiliations de l'espèce avec la trame. La méthode initiale a cependant l'avantage considérable de prendre en compte toutes les espèces rares et non seulement celles qui ont été sélectionnées comme indicatrice. Cependant, il semblerait tout de même qu'il y ait des surestimations pour la trame bocagère notamment. De ce fait, il serait préférable d'identifier une méthode alternative basée sur les espèces indicatrices n'étant pas contrainte par le nombre.

Idéalement, il aurait été préférable d'ajuster la taille des zones cœurs de biodiversité en fonction de celle des domaines vitaux spécifiques à chaque espèce indicatrice et de supprimer les ensembles de carrés kilométriques n'atteignant pas cette superficie. Cependant, après consultation de la littérature (Harrison et al., 1959; Smedshaug et al., 2002; Wernham et al., 2002; Conway et al., 2009) et l'avis d'experts, il semblerait qu'une taille minimale d'un kilomètre carré soit suffisante pour la grande majorité des espèces. Cela a été également compliqué dans le temps imparti de parvenir à récolter cette information pour l'ensemble des espèces indicatrices. Il en est de même pour la distance de dispersion qui aurait été forte utile lors de l'élaboration des corridors.

Bien entendu, toutes les zones centrales déterminées par trames n'ont pas toutes la même importance. Elle peut être individuellement estimée via le delta métrique appelé dPC. L'évaluation peut se jouer alors sur trois aspects différents, la capacité d'accueil de la tache, sa participation à la connectivité globale et sa capacité de réservoir biologique. Le premier dans ce cas-ci se résume uniquement à la taille de la tache. Le second mesure sa participation à lier d'autres taches entre elles ce qui est très important dans l'évaluation d'un réseau quelconque. Finalement, la dernière composante mesure la capacité de réservoir en espèce de la tache, prenant en compte les taches voisines auxquelles elle est connectée. Sur l'entièreté des cartes présentées en annexe 6 ou au sein des figures 15 et 16, il s'agit du dPC connector (2<sup>ème</sup> dPC) mesuré avec la distance de dispersion de 10km qui est représenté. Cette distance a été choisie pour cette représentation, car il s'agit de la distance la plus élevée permettant ainsi d'avoir également une évaluation des taches assez éloignées. Cependant, il faut juger l'importance d'une tache sur base de l'entièreté des indices et ceci pour les trois distances de dispersion. En effet, une tache pourrait gagner en importance à mesure où la distance de dispersion des espèces diminue. C'est pourquoi un ordre de priorité est proposé qui est le suivant : favoriser ceux ayant une bonne valeur pour au moins deux distances différentes suivi de ceux ayant de bons résultats avec une distance de 2km, puis 10 et enfin la distance intermédiaire. De cette façon, les zones cœurs isolées seront moins favorisées que leurs homologues connectées entre elles ce qui respecte parfaitement l'idée d'un réseau écologique. D'un point de vue plus global, l'accent pourrait également être mis sur les zones cœurs communes entre les différentes trames. De cette manière, cela pourrait mieux justifier l'échelle kilométrique et assurer une utilité plus complète en matière de maintien de la biodiversité.

Sur tous les pixels de 1km<sup>2</sup> ayant été sélectionnés comme zone cœur, le nombre maximal de trames regroupées au même endroit est de 5 et sont du nombre de 24. La plupart d'entre eux se trouvent dans le Condroz alors que la Fagne-Famenne et le NSSM en comptent chacun deux. Il s'agit le plus fréquemment de milieux forestiers proches de cours d'eau. Grâce à la sélection par région biogéographique des zones cœurs, la distribution du réseau global est assez homogène sur toute la Région Wallonne. Étonnement, il ne prend en compte qu'une faible proportion surfacique du Parc naturel de Gaume. Ceci est majoritairement dû à une moyenne de présence d'espèces estimée assez élevée en Lorraine que celle du parc.

Quelques particularités peuvent être observées au sein des différentes régions. Pour commencer, les trames des milieux agricoles et des bocages sont absentes de la Lorraine. Ensuite, le réseau des milieux ouverts est assez particulier comparé aux autres. En effet, il possède une superficie bien inférieure et ses zones cœurs sont principalement cantonnées en Lorraine et en Ardenne. Il est très probable qu'une erreur ait été commise lors de la sélection des seuils pour cette trame concernant l'indice alpha modélisé. Ceci aurait eu pour conséquence de limiter fortement sa représentation dans les régions où la surface d'occupation du sol de cette trame est déjà assez faible.

Il n'est pas étonnant en prenant compte l'état de la Région wallonne que la trame des prairies et des milieux agricoles en général aient les plus grandes surfaces. Cette première trame ne compte pas moins de 1397 km<sup>2</sup> de zones cœurs. En ne prenant maintenant en compte que des trames secondaires, celle des lisières affiche la surface la plus importante. Ceci est assez logique étant donné l'échelle de travail et qu'elle regroupe les lisières en bord des deux types de massifs forestiers. Par ailleurs, il s'agit bien de la trame dont la mesure de la surface totale à la moins de sens de cette manière. Par contre, il est intéressant de voir que la superficie du réseau forestier global est inférieure à celui des forêts feuillues, mais pas des forêts résineuses. La trame des milieux agricoles se positionne à peu près de la même façon par rapport à ses trames secondaires. Ainsi, les prairies et les forêts feuillues seraient des milieux accueillant une meilleure diversité moyenne en oiseau que leurs trames homologues.

Lors de la comparaison deux à deux des différents réseaux, des phénomènes assez logiques apparaissent. Par exemple, plus de 50% du réseau des lisières se partage avec les trois trames forestières. La comparaison inverse donne la même chose sauf pour les forêts feuillues qui ne partagent que 48 % de leur réseau avec celui des lisières. La distinction entre les zones de pâtures et de culture se justifie bien étant donné que leurs réseaux respectifs partagent peu de zones cœurs. Également, un peu moins de 40% des zones ouvertes désignées le sont également pour la trame résineuse avec laquelle elle possède sa deuxième plus haute proportion de surface partagée. Cela pourrait s'expliquer en partie car de nombreux habitats classés au sein des milieux ouverts naturels sont encore en cours de restauration impliquant une certaine proximité avec les forêts résineuses. De ce fait, certaines taches forestières destinées à être restaurées en milieux ouverts pourraient être tout de même relativement importantes pour l'avifaune affiliée aux Gymnospermes. Pour finir, il est assez encourageant de voir que les milieux ouverts partagent plus de 50% de leurs zones centrales de biodiversité avec la trame des lisières. Cependant, cela indique tout de même que la résolution utilisée brouille la transition entre les milieux forestiers et ouverts.

#### 4.4.3. Délimitations des corridors

Le logiciel Graphab a permis de générer l'entière des liens pouvant exister entre les différentes taches tout en évitant de relier des taches géographiquement opposées via la création d'un graphe planaire. Lors du calcul de différentes métriques et de la génération des graphes par distances de dispersion, les distances intra-taches ne sont pas prises en compte, car l'hypothèse est que le déplacement au sein de la zone d'habitat peut être considéré comme « gratuit ». De cette manière, le déplacement à l'intérieur d'une tache n'est pas pris en compte dans la distance inter-tache.

Parmi les trois composantes du dPC calculées, seul le dPC connector peut être interprété et informe sur l'importance de chaque lien. C'est pourquoi ceux ayant une valeur de zéros peuvent être jugés inutiles et supprimés. Pour certains réseaux, cela crée une sélection raisonnable et suffisante, mais pour la plupart, le nombre de liens restants est toujours assez important. Malgré tout, trouver un seuil convenable permettant d'obtenir un nombre de liens correct sans le diminuer de trop semble assez complexe. De ce fait, il est clairement préférable d'estimer les corridors fonctionnels via une couche raster générée à partir des points centraux de ces liens afin d'obtenir la représentation cartographique de la connectivité elle-même. Procéder de cette manière est plus avantageux que de résumer le déplacement probable d'une espèce d'oiseau à une simple ligne. Ces couches rasters mettent en avant les zones où les densités de connectivités sont les plus importantes impliquant une nécessité d'y protéger les éléments du maillage écologique.

La dernière couche raster créée représente quant à elle l'importance d'une zone par rapport à la densité en corridor, mais également et surtout par rapport à son importance dans la connectivité totale du réseau. Concrètement, la couche non pondérée représente les corridors pouvant exister entre les différentes taches et la seconde, les zones à prioriser pour le maintien d'une bonne connexion au sein du réseau.

#### 4.5. *Analyse des réseaux*

Pour commencer, les indices globaux mesurés pour chaque réseau permettent de les évaluer par rapport à leur connectivité selon la disposition des zones cœurs et le nombre de liens existants. La PC comme son nom l'indique, vaut la probabilité que deux individus mis au hasard au sein du réseau soient capables de se rejoindre. Il est donc normal qu'au plus la distance de dispersion considérée diminue, au plus cette probabilité se réduit. Cependant, toutes les valeurs obtenues sont assez basses comparées à la valeur maximale. Cela s'explique par l'étendue assez importante de l'aire d'étude qui favorise la création de groupe de taches isolées diminuant la connectivité globale. De cette manière, trois cas de figure apparaissent ici concernant cet indice. Le premier caractérise les réseaux ayant une faible superficie diminuant fortement leur PC s'ils s'étendent sur toute la région comme la trame des milieux ouverts et dans une moindre mesure, celle des cultures. Le deuxième cas apparaît avec les trames ayant une superficie relativement grande comme celle des forêts feuillues ou les prairies montrant les PC les plus élevées. Finalement, le dernier correspond aux trames ayant également une surface importante, mais dont la valeur d'indice reste plus faible que la moyenne. L'explication est la même que précédemment où la création de groupes isolés réduit cette connectivité globale. Par exemple, bien qu'elle possède une surface de plus de 800km<sup>2</sup>, la trame des forêts résineuses a une valeur d'indice assez basse. Malgré tout, un réseau configuré de cette manière avec de nombreux réseaux « secondaires » isolés n'est pas forcément mauvais tant que ces derniers restent relativement grands et que les taches qui les composent soient bien connectées.



Le cas le plus dommageable pourrait bien être celui des milieux ouverts, car les zones cœurs au nord de la Wallonie sont fortement isolées.

En comparant les valeurs des PC de la même trame, il est possible d'identifier les réseaux les plus sensibles à la variation de la distance de dispersion des espèces. La trame des prairies gagne beaucoup en connectivité lorsqu'elle passe à la plus haute catégorie de distance ce qui n'est pas le cas pour les milieux ouverts. En effet, cette trame a de base, une valeur d'indice faible, car elle est constituée de nombreux sous-groupes distants de plus de 10km l'un de l'autre. Ces zones cœurs se concentrent principalement en Lorraine, en Ardenne centrale et orientale et légèrement en région limoneuse et en Fagne-Famenne. L'aspect positif est que ces sous réseaux possèdent une bonne intraconnexion ce qui réduit l'influence de la distance de dispersion des espèces dans leur capacité de déplacement au sein du réseau. Pour la trame précédente, même si elle enregistre de grandes variations, l'indice reste supérieur à la moyenne obtenue par distance. Il en est de même pour les trois trames suivantes qui enregistrent le plus de variations avec le changement de distance.

Comme dit précédemment, la PC est assez sensible aux sous-réseaux. Grâce à l'indice d'Harary, une mesure directe du rapport entre le nombre de liens séparant les différentes zones cœurs est faite, permettant la caractérisation de la connectivité globale du réseau via uniquement le nombre de corridors. Il est donc une fois de plus normal que les réseaux construits via la distance de dispersion moyenne de 10km obtiennent les meilleurs scores. Également, ceux ayant une superficie assez importante comme la trame des prairies ou des forêts ont de meilleurs résultats étant donné que le nombre de corridors est directement lié au nombre de zones cœurs. Comme il est fort variable en fonction des trames, il a fallu mesurer l'indice de manière relative à la valeur maximale que celui-ci peut prendre. De cette manière, une comparaison directe entre les trames est possible. Par exemple, la trame des forêts feuillues possède une valeur d'indice relativement faible alors qu'elle possède tout de même une superficie importante. Cependant, via son indice relatif assez élevé, il peut en être déduit que les zones cœurs des forêts feuillues sont moins nombreuses, mais de superficie plus importante. De plus, la valeur brute minimale est atteinte par la trame des milieux ouverts, suivis par celle des milieux aquatiques. Or, lorsqu'elles sont mises en relation avec leur valeur maximale, la trame des milieux ouverts passe dans les premières. Ceci signifie que relativement à son nombre de zones cœurs, elle possède un nombre correct de corridors impliquant tout de même une bonne connectivité contrairement à ce qu'indique l'indice précédent. Au contraire, les milieux aquatiques se positionnent toujours relativement bas validant le fait qu'il n'y ait pas assez de corridors comparés aux nombres de zones cœurs. D'un point de vue général, la valeur de cet indice n'est d'ailleurs pas très bonne, car elle reste autour de 10%. Ceci pourrait induire des effets négatifs pouvant entraver le déplacement des oiseaux au sein d'un réseau. De plus, cet indice possède comme défaut de prendre également en compte les corridors supprimés en raison de leur inutilité mesurée via le dPC. Ceci a dû surestimer la valeur des deux indices d'Harary ce qui signifie que la connectivité réelle des réseaux est moins bonne que celle mesurée. Une étude plus approfondie pourrait être faite afin d'améliorer ceci en sélectionnant de nouvelles zones cœurs spécifiquement pour augmenter la connectivité. Pour terminer, de la même manière que pour la PC, certains réseaux semblent bien plus sensibles à la variation de la distance de dispersion. La trame des cultures enregistre une très grande variation impliquant une dépendance importante du nombre de liens existants et la distance de dispersion des espèces. Ces deux indices offrent des résultats assez contrastés l'un de l'autre, car même s'ils caractérisent tous les deux la connectivité d'un réseau, ils ne sont pas sensibles aux mêmes phénomènes. En effet la PC est bien plus sensible à la présence de sous-réseaux alors que l'autre indice, non.

D'un autre côté, l'existence de quatre liens entre chaque zone cœur n'est pas fondamentale et ne limite la mesure de la connectivité. C'est pourquoi ces deux indices doivent être utilisés ensemble pour bien décrire la connectivité et la qualité d'un réseau écologique.

#### 4.6. *Analyses complémentaires*

Pour commencer, trois indices ont été utilisés afin de démontrer la pertinence des réseaux écologiques. Dans ce but, deux points importants doivent être respectés. Le premier stipule que les réseaux devraient idéalement posséder une meilleure diversité que celle de la matrice paysagère. Le second voudrait qu'idéalement, les zones les plus populaires chez les ornithologistes soient incorporées dans un réseau. Pour se faire, les deux indices de biodiversité disponibles sont à nouveau utilisés. Il semblerait que pour la majorité des trames, les moyennes de ces indices augmentent que ce soit par rapport à la moyenne mesurée sur toute une région ou uniquement mesurée dans les carrés kilométriques hors réseau. Comme un des principaux filtres de sélection se base sur la diversité alpha modélisée, sa moyenne au sein des zones cœurs augmente forcément. Les trames associées aux milieux agricoles offrent généralement les valeurs d'indice les plus faibles particulièrement pour les cultures. Il en est de même pour l'indice alpha non modélisé. Celui-ci est moins dépendant de la méthode de sélection et est très souvent plus faible que les deux moyennes globales pour ces trames. Cependant, la variance dans la distribution de cet indice ne permet pas de conclure sur de réelles différences. De plus, il est probable que la faible diversité de ces milieux soit intrinsèque à leurs conditions. C'est pourquoi l'interaction entre l'occupation du sol et la valeur de ces indices a été mesurée via la corrélation entre ces variables. Les cultures possèdent une corrélation négative pour les trois indices dans chaque région sauf en Ardenne pour l'indice alpha pondéré. Ces valeurs sont certes faibles, mais peuvent tout de même expliquer cette diversité assez faible. Par ailleurs, les variances du premier indice sont très élevées compliquant la comparaison des moyennes qui n'aboutit pas sur des différences significatives. L'analyse du  $\chi^2$  concernant la popularité des sites a permis de conclure sur une proportion significativement plus importante de sites populaires au sein du réseau global. Enfin les proportions attendues dans le tableau 23 sont bien plus faibles que celles réellement obtenues pour les sites populaires au sein d'un réseau.

Après l'analyse de la connectivité et de la diversité spécifique au sein de chaque réseau, une nouvelle approche doit être entreprise. Celle-ci consiste à réaliser une évaluation de la précision globale de chaque réseau par rapport à la trame qu'ils représentent. Dans un premier temps, il y a la fraction de la trame protégée qui renseigne concrètement sur l'étendue de celle-ci qui est couverte par son réseau. Instinctivement, au plus celle-ci est grande, au mieux c'est sauf que cet indice n'est pas directement indicateur de la qualité d'un réseau. En effet, elle dépend forcément de la surface totale de la trame et pour celles qui sont le plus répandues, une faible valeur peut être attendue sans que cela soit problématique. Généralement, ce sont les milieux ouverts là où ils sont représentés et les milieux aquatiques qui montrent les plus grandes valeurs de cette mesure. Cela s'explique par la faible superficie de ces deux trames. Par ailleurs, une redésignation des zones cœurs au nord du sillon et dans le Condroz devrait être envisagée pour cette première trame. En effet, elle y est très peu représentée faisant diminuer cette proportion respectivement à 2,6 et 0,5%. En Lorraine également, les trames forestières y sont peu représentées. D'un point de vue général, il est important que la valeur de cet indice se situe entre 10 et 20%.

Pour l'évaluation des réseaux, l'indice de précision semble plus pertinent. Il renseigne sur la capacité d'un réseau à couvrir sa propre trame écologique et à en être spécifique. Le fait de séparer à nouveau le travail par région permet d'identifier celles pouvant également nécessiter la redéfinition du réseau d'une trame. Cependant, cette mesure est tout de même assez dépendante de la surface moyenne au sein d'un carré kilométrique de la trame. En effet, si celle-ci est basse, il faut s'attendre à ce que la précision le soit également. Les milieux aquatiques en sont un bon exemple, car les eaux de surface ne représentent généralement qu'une faible proportion d'un carré kilométrique. La trame des forêts résineuses possède également souvent une faible précision même en Ardenne où elle n'est à peine plus élevée que 30%. Ceci envisagerait le fait qu'au final, peu de surfaces résineuses sont présentes dans son réseau. Cependant, il est important d'analyser également le dernier indice mesuré qui estime la différence proportionnelle de précision entre un réseau et la matrice paysagère. Celle-ci est assez importante que les résineux tout comme les milieux ouverts et aquatiques. Ainsi, la faible valeur de l'indice de précision de ces trames peut donc s'expliquer par l'échelle qui est globalement trop grande pour obtenir une valeur plus élevée. Pourtant, l'échelle globale ne semble pas parvenir à expliquer les valeurs assez basses pour les bocages au nord du sillon et dans le Condroz. En effet, les valeurs du dernier indice  $\gamma$  sont négatives indiquant une diminution de surface de la trame lorsqu'uniquement les zones cœurs sont considérées. Cela signifierait que le réseau n'y offre pas une bonne représentation de sa trame et certaines modifications devraient donc y être faites.

Tous ces processus et manipulations mis en place devraient en théorie offrir une cartographie relativement précise selon l'échelle utilisée des sites les plus importants pour la nidification des espèces d'oiseaux nichant en Région Wallonne. Ce qui voudrait dire qu'une grande proportion de ces sites ainsi désignés devrait se trouver d'ores et déjà sous réglementations. Or, ça n'est visiblement pas autant le cas que cela devrait être en vue du tableau 25. En effet, seulement 1,63% du réseau global est déjà protégé par la LCN et 8% de sa surface est reconnue comme SGIB. Ainsi, un peu plus de 47% des surfaces sous protection de la LCN ont été sélectionnés dans un réseau. Ce pourcentage diminue à 40% pour les SGIB. Cela indiquerait que d'un point de vue de l'avifaune, plus de la moitié des sites remarquables ne le sont pas pour elle, plus de 90% des sites importants ne sont pas encore reconnus et 98% ne sont pas protégés. En vue de la très faible proportion des terres wallonnes qui sont efficacement protégées, ce pourcentage n'est pas étonnant. Par contre, en suivant les seuils indiqués dans le tableau 13, seulement 43% des sites ayant une biodiversité en oiseau élevée et reconnue sont pris au sein d'un réseau ce qui vaut 17% du réseau global. Cependant, cet indice n'est mesuré uniquement sur base des observations des espèces rares. De plus, lors de l'analyse de l'indice alpha observé, une augmentation de celui-ci peut être observée lorsque seules les zones cœurs sont prises en compte mais reste tout de même non significative.

## V. Conclusion

### 5.1. Méthodes de modélisation

Théoriquement, l'algorithme de MaxEnt se prête bien mieux à la modélisation de données de présence uniquement, comparé à GAM. En effet, ce dernier demande nécessairement des absences sans quoi, son efficacité de prédiction diminue grandement. Ainsi, MaxEnt serait donc conseillé dans ce cas d'étude-ci étant donné la nature des données à modéliser. Cependant, même si MaxEnt s'en sort très bien avec, le biais d'échantillonnage lié à ces données d'observation dites « civiles » est tout de même important et cause une perte de qualité des modèles s'il n'est pas contrôlé. Le biais principal provient de ce manque de points d'absences et par la nécessité de sélection de points aléatoires résultant sur un plus haut taux de fausses absences. Pour pallier à cela, MaxEnt préconise l'utilisation d'un grand nombre de points de pseudo-absence afin d'également augmenter le nombre de vraies absences d'où la méthode par défaut utilisant 10 000 points. Cependant, il est possible de réduire ce taux en les sélectionnant de façon plus scrupuleuse. Ceci peut être réalisé en estimant les efforts d'échantillonnage au sein de chaque unité afin que l'absence d'une espèce puisse être considérée comme vraie. Grâce aux deux méthodes testées pour y parvenir, non seulement la sélection des points d'absence est améliorée mais il est également possible de tester l'algorithme de GAM. Malheureusement, l'utilisation des données de présence d'autres espèces a des effets négatifs qui combinés avec la réduction du nombre de points d'absence considérés, l'obtention de modèle trop généreux est trop fréquente. Même si l'idée initiale est encourageante, il semblerait en vue des résultats de l'analyse de sensibilité que la méthode par défaut de MaxEnt soit la plus simple et la plus efficace pour les données disponibles. Une étude plus approfondie sur une amélioration de l'utilisation de MaxEnt et même de GAM pourrait être envisagée et permettrait sans doute la réalisation de modèles encore meilleurs.

### 5.2. Modélisation

Pour la réalisation de ce travail, la niche écologique potentielle de 44 espèces classées comme rares a été modélisée. Neuf d'entre elles ont été mises sur liste rouge et quatre parmi celle-là possèdent de très mauvaises valeurs d'indices. Ces derniers ont permis dans un premier temps de sélectionner la meilleure méthode de modélisation même si l'analyse cartographique a eu le plus de poids. Ensuite, ils ont permis de sélectionner les meilleurs sets de variables à être utilisés dans le modèle d'une espèce. La couche écotope a servi de base fondamentale à cette modélisation en offrant toutes les variables environnementales nécessaires à cette modélisation de l'avifaune. Également, elles ont été mesurées à nouveau de façon spécifique aux carrés kilométriques de l'Atlas européen des oiseaux.

De tous les indices utilisés, le User accuracy a toujours été prioritaire étant donné qu'il reflète l'erreur de commission qui est à minimiser lors de la désignation de zones de protection. Cependant, il a peut-être eu une trop grande importance spécialement lors de la création de la liste rouge. Quoi qu'il en soit, ils ont tous servi à leur manière à la désignation finale des meilleurs modèles et à leur évaluation. Concernant ce dernier point, mis à part l'indice de corrélation de Mathew, tous ces indices sont dépendant de la prévalence ce qui a demandé des ajustements pas forcément pratiques lors des comparaisons entre deux espèces.

### 5.3. *Construction des réseaux écologiques*

L'utilisation des analyses de la redondance via les données d'abondance d'une part et des probabilités de MaxEnt de l'autre a facilement permis d'obtenir une liste d'espèces indicatrices par trame écologique. Cependant, elles doivent être exécutées avec les bonnes variables environnementales qui décrivent bien chaque trame. Une fois de plus, les variables écotopes d'occupation du sol sont les meilleures candidates à remplir ce rôle. La trame des bocages et lisières a dû être séparée car aucune variable ne la représente. Pour les bocages, une carte spécifique ayant été construite par interprétation spatiale a pu être utilisée. Les lisières quant à elles n'ont pas eu de carte les représentant précisément. L'alternative a été d'utiliser la couche écotope représentant les milieux forestiers perturbés où les espèces qui y sont affiliées peuvent être considérées comme l'étant également aux lisières donnant à première vue des résultats corrects.

La sélection des zones cœurs potentielles s'est beaucoup basée sur la détermination de seuils classifiant les carrés kilométriques. Cette façon de faire risque même avec les méthodes utilisées, d'apporter une certaine subjectivité dans la sélection. Cependant, il semblerait que cela soit indispensable afin de désigner efficacement des zones d'intérêt sur un grand territoire. La méthode utilisée n'est certainement pas sans défauts mais les résultats finaux sont tout de même encourageants. En effet, l'évaluation des différents réseaux montre une augmentation de la diversité observée et modélisée pour la grande majorité des trames dans chaque région. De plus, il semblerait également que le nombre d'observations moyen soit également à la hausse. Cela voudrait dire que globalement, cette méthode permet la sélection des spots populaires chez les ornithologues. Cependant, pour ce premier et dernier indice, les variances sont très élevées ne rendant pas cette supériorité réellement significative. Des tests statistiques plus poussés pourraient être envisagés pour confirmer cette supériorité. Pour finir, chaque réseau représente globalement bien sa trame dans chaque région. Cependant, les trames des milieux ouverts et des bocages devraient être retravaillées spécialement au nord du sillon et dans le Condroz tout comme les trames forestières en Lorraine.

En conclusion, la méthode globale qui a été utilisée offre des résultats satisfaisants et est assez prometteuse. Cependant, deux points majeurs à améliorer sont assez évidents à commencer par la modélisation des espèces où des approches plus fortes qui minimisent le biais d'échantillonnages devraient être utilisées. Finalement, l'utilisation de l'indice alpha pondéré devrait pouvoir être évitée en trouvant une méthode basée directement sur le nombre d'espèces indicatrices qui affinerait mieux les réseaux et en plus, trouverait une alternative pour les trames ayant peu d'espèces indicatrices.

## VI. Bibliographie

- A.Hanley J. & J.McNeil B., 1982. The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. *Radiology*, **143**(5), 29–36.
- A.Swets J., 1988. Measuring the Accuracy of Diagnostic Information. *Science*, **240**, 1286–1293.
- Akaike H., 1973. Maximum likelihood identification of gaussian autoregressive moving average models. *Biometrika*, **60**(2), 255–265.
- ALBERT C.H. & CHAURAND J., 2018. Comment choisir les espèces pour identifier des réseaux écologiques cohérents entre les niveaux administratifs et les niveaux biologiques ? *Sci. Eaux Territ.* **25**, 26–31.
- Amsallem J., Bonneville M. & Deshayes M., 2013. Analyse comparative de méthodes d'élaboration de trames vertes et bleues nationales et régionales. *Sci. Eaux Territ.*, **3**, 40–45.
- Anderson R.P., Lew D. & Peterson A.T., 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Commun. Theor. Phys.*, **162**, 211–232.
- Anderson R.P. & Raza A., 2010. The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: Preliminary tests with montane rodents (genus *Nephelomys*) in Venezuela. *J. Biogeogr.*, **37**(7), 1378–1393.
- Austin M., 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecol. Modell.* **200**(1–2), 1–19.
- Austin M.P. & Van Niel K.P., 2011. Improving species distribution models for climate change studies: Variable selection and scale. *J. Biogeogr.*, **38**(1), 1–8.
- Aves.be, 2015. L'Atlas européen des oiseaux nicheurs: en Wallonie.  
<https://www.aves.be/index.php?id=3410>, (16/05/2020).
- Avon C. & Bergès L., 2016. Prioritization of habitat patches for landscape connectivity conservation differs between least-cost and resistance distances. *Landsc. Ecol.*, **31**(7), 1551–1565.
- Ayram C.A.C. et al., 2014. Identifying potential conservation areas in the Cuitzeo Lake basin, Mexico by multitemporal analysis of landscape connectivity. *J. Nat. Conserv.*, **22**(5), 424–435.
- Baker S.G. & Pinsky P.F., 2001. A proposed design and analysis for comparing digital and analog mammography: Special receiver operating characteristic methods for cancer screening. *J. Am. Stat. Assoc.*, **96**(454), 421–428.
- Barbet-Massin M. et al., 2012. Selecting pseudo-absences for species distribution models: How, where and how many? *Meth. Ecol. Evol.*, **3**(2), 327–338.

- Beier P., Majka D.R. & Spencer W.D., 2008. Forks in the road: Choices in procedures for designing wildland linkages. *Conserv. Biol.*, **22**(4), 836–851.
- Benesty J. et al., 2009. Noise Reduction in Speech Processing. Pearson Correlation Coefficient *Noise Reduct. speech*, **2**, 229.
- Bernier A. & Théau J., 2013. Modélisation de réseaux écologiques et impacts des choix méthodologiques sur leur configuration spatiale : analyse de cas en Estrie (Québec, Canada), *VertigO*, **13**(2), 1–38.
- Biodiversité.wallonie.be, 2020. Réserves naturelles & Co.  
<http://biodiversite.wallonie.be/fr/reserves-naturelles-co.html?IDC=825>, (15/05/2020).
- Blagus R. & Lusa L., 2010. Class prediction for high-dimensional class-imbalanced data. *BMC Bioinfo*, **11**, 523.
- Borcard D., Gillet F. & Legendre P., 2008. *Numerical Ecology with R*, Applied Spatial Data Analysis with R. Gentleman, R., Hornik, K., Parmigiani, G.G. eds., Springer Science+Buisness Media, New York, 1–306.
- Bourdouxhe A. et al., 2020. Comparison of methods to model species habitat networks for decision-making in nature conservation : the case of the wildcat in Southern Belgium. 1–37 (à paraître).
- Broennimann O. & Guisan A., 2008. Predicting current and future biological invasions: Both native and invaded ranges matter. *Biol. Lett.*, **4**(5), 585–589.
- Brogna D. et al., 2018. Forest cover correlates with good biological water quality. Insights from a regional study (Wallonia, Belgium). *J. Environ. Manage.* **211**, 9–21.
- Bunn A.G., Urban D.L. & Keitt T.H., 2000. Landscape connectivity: A conservation application of graph theory. *J. Environ. Manage.*, **59**(4), 265–278.
- Camargo J.L. & Kapos V., 1995. Complex edge effects on soil moisture and microclimate in central Amazonian forest. *J. Trop. Ecol.*, **11**(2), 205–221.
- Chetkiewicz C.-L.B., St. Clair C.C. & Boyce M.S., 2006. Corridors for Conservation: Integrating Pattern and Process. *Annu. Rev. Ecol. Evol. Syst.*, **37**(1), 317–342.
- Chicco D. & Jurman G., 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, **21**(1), 1–13.
- Chmura D. & Edyta S., 2006. relation between invasive plant and species richness of forest floor vegetation : a study of *Impatiens Parviflora* DC . *Pol J. Ecol.*, **54**(3), 417–428.
- Commission Européenne, 2013. Programme d'action général de l'Union pour l'environnement à l'horizon 2020 «Bien vivre, dans les limites de notre planète». *J. Off. l'Union Eur.* 171–200.
- Conway G., Wotton S., Henderson I., Eaton M., Drewitt A. & Spencer J., 2009. The status of breeding woodlarks *Lullula arborea* in Britain in 2006. *Bird Study*, **56**(3), 310–325.

- Council of Europe, 1996. Pan-European biological and landscape diversity strategy, Nature and environment, Strasbourg, 50.
- Council of Europe, 2009. “The Emerald Network — Network of areas of special conservation interest,” Strasbourg.
- De Blust G., Kuijken E. & Paelinx D., 1992. De groene hoofdstructuur van Vlaanderen. *Ruimtelijke Plan.* **30**, 1–30.
- Delescaillle L., 1993. Le maillage écologique et l’espace rural. *Ann. Gembloux* **99**, 61–69.
- Derouaux A. & Paquet J., 2018. L’évolution préoccupante des populations d’oiseaux nicheurs en Wallonie : 28 ans de surveillance de l’avifaune commune. *Aves*, **55**(1), 1–31.
- Devillers P. et al., 2003. Évaluation de la gestion d’un réseau de zones protégées. *Les Nat. belges* **84**(1), 15–24.
- Dobson A.J., 2002. An Introduction To Generalized Linear Models Second Edition, Chapman & Hall., New York.
- Dormann C.F. et al., 2013. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography (Cop.)*, **36**(1), 027–046.
- Duflot R. et al., 2018. Combining habitat suitability models and spatial graphs for more effective landscape conservation planning: An applied methodological framework and a species case study. *J. Nat. Conserv.* **46**, 38–47.
- Dufrêne M., 2004. Réseau écologique - Structure écologique principale: Concepts, structure et stratégie d’élaboration. Version 1.0, Gembloux, 1–34.
- Dufrêne M., 2005. Le réseau Natura2000 en Région wallonne : stratégies , réseau écologique et mise en oeuvre, Chaire Tractebel, Gembloux, 167–185.
- Dufrêne M. & Delascaillle L.-M., 2005. *La typologie WalEUNIS des biotopes wallons*.
- Dufrêne M. & Legendre P., 1997. Species assemblages and indicator species\_the need for flexible asymmetrical approach. *Ecol. Monogr.* **67**(3), 345–366.
- Dufrene M., Legendre P. & Croix P., 1991. Geographic structure and potential ecological factors in Belgium **18**(3), 257–266. <http://www.jstor.com/stable/2845396> (04/05/20).
- Dufrêne M. et al., 2019. Guide méthodologique provisoire de la cartographie du réseau écologique à l’échelle locale, Gembloux, 1–23.
- Dufrêne M. et al., 2006. A research project to study patterns, roles and determinants of wood-dependent species diversity in Belgian deciduous forests (Xylobios). *Belgian Sci. Policy.*, 5-118.
- Dupont S. et al., 2011. Long-distance edge effects in a pine forest with a deep and sparse trunk space: In situ and numerical experiments. *Agric. For. Meteorol.*, **151**(3), 328–344.



- Elith J. & Leathwick J.R., 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annu. Rev. Ecol. Evol. Syst.*, **40**(1), 677–697.
- Elith J. et al., 2011. A statistical explanation of MaxEnt for ecologists. *Divers. Distrib.*, **17**(1), 43–57.
- Evans D., 2006. The habitats of the European union habitats directive. *Biol. Environ.*, **106B**(3), 167–173.
- Evans D., 2012. Building the European Union's Natura 2000 network. *Nat. Conserv.*, **1**, 11–26.
- Fahrig L., 2003. Effects of Habitat Fragmentation on Biodiversity. *Annu. Rev. Ecol., Evol. Syst.* **34**, 487–515.
- Fawcett T., 2006. An introduction to ROC analysis. *Pattern Recognit. Lett.*, **27**(8), 861–874.
- Fielding A.H. & Bell J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.*, **24**(1), 38–49.
- Foltête J.C., Clauzel C. & Vuidel G., 2012. A software tool dedicated to the modelling of landscape networks. *Environ. Model. Softw.*, **38**, 316–327.
- Forman R.T.T. & Godron M., 1986. *Landscape ecology*. Wiley ed., John Wiley and Sons, New York, 619.
- Fourcade Y. et al., 2014. Mapping species distributions with MAXENT using a geographically biased sample of presence data: A performance assessment of methods for correcting sampling bias. *PLoS One*, **9**(5), 1–13.
- Fung T. & Ledrew E., 1988. The determination of optimal threshold levels for change detection using various accuracy indices. *Photogramm. Eng. Remote Sens.*, **54**(10), 1449–1454.
- Girardet X. & Vuidel G., 2017. Graphab 2.2. Manuel d'utilisation, 2-36.
- Gonzalez S.C., Soto-Centeno J.A. & Reed D.L., 2011. Population distribution models: Species distributions are better modeled using biologically relevant data partitions. *BMC Ecol.*, **11**.
- Green H.. D.R., 1979. *Sampling Design and Statistical Methods for Environmental Biologists*, John Wilhey & Sons, Ontario.
- Green infrastructure, 2007. Toward a Green Infrastructure For Europe: Developing new concepts for integration of Natura 2000 network into a broader countryside.
- Guisan A., Edwards T.C.J. & Hastie T., 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecol. Modell.*, **157**, 89–100.
- Guisan A., Thuiller W. & Niklaus E. Z., 2017. *Habitat Suitability and Distribution Models*, Cambridge University Press, New York.

- Hansson L., Solbreck C. & Söderström L., 1992. *The Ecology of Dispersal in Relation to Conservation*, Ecological Principles of Nature Conservation, Springer Boston MA, Boston, 162–200.
- Harezlak J., Ruppert D. & Wand M.P., 2018. *Generalized Additive Models*, Semiparametric Regression with R. Use R!, Springer, New York, New York.
- Harrell F.E.J., 2001. *Regression Modeling Strategies: With Applications to Linear Models, Logistic And Ordinal Regression, and Survival Analysis*, Springer, New York, New York, 581.
- Harrison C.J.O. & Forster J., 1959. Woodlark territories. *Bird Study*, **6**(2), 60–68.
- Hastie T.J. & Tibshirani R.J., 1990. *Generalized Additive Models*, Chapman & Hall., Toronto, 308.
- Hengl T. et al., 2009. Spatial prediction of species' distributions from occurrence-only records: combining point pattern analysis, ENFA and regression-kriging. *Ecol. Modell.*, **220**(24), 3499–3511.
- Hilty J.A. & Merenlender A.M., 2004. Use of Riparian Corridors and Vineyards by Mammalian Predators in Northern California. *Conserv. Biol.*, **18**(1), 126–135.
- Huang J. & Ling C.X., 2005. Using AUC and accuracy in evaluating learning algorithms. *IEEE Trans. Knowl. Data Eng.*, **17**(3), 299–310.
- Hutchinson G.E., 1957. Concluding Remarks. *Cold Spring Harb. Symp. Quant. Biol.*, **22**(0), 415–427.
- Iturbide M. et al., 2015. A framework for species distribution modelling with improved pseudo-absence generation. *Ecol. Modell.*, **312**, 166–174.
- Jackson S.T. & Overpeck J.T., 2000. Responses of plant populations and communities to environmental changes of the late Quaternary. *Paleobiology*, **26**(sp4), 194–220.
- Jacob, J-P. et al., 2010. *Atlas des oiseaux nicheurs de Wallonie 2001-2007*, Aves et Département de l'Etude du Milieu Naturel et Agricole, Gembloux.
- Jaeger J.A.G. et al., 2016. *Landscape fragmentation in Europe*, European Landscape Dynamics: CORINE Land Cover Data, 157–198.
- Jiménez-Valverde A. & Lobo J.M., 2007. Threshold criteria for conversion of probability of species presence to either-or presence-absence. *Acta Oecologica* **31**(3), 361–369.
- Johnston A., Fink D., Hochachka W.M. & Kelling S., 2018. Estimates of observer expertise improve species distributions from citizen science data. *Methods Ecol. Evol.* **9**, 88–97.
- Johnston A. et al., 2020. Estimating species distributions from spatially biased citizen science data. *Ecol. Modell.*, **422**.
- Jongman R.H.G., 2002. Homogenisation and fragmentation of the European landscape: Ecological consequences and solutions. *Landsc. Urban Plan.*, **58**(2–4), 211–221.

- Jongman R.H.G. et al., 2011. The pan European ecological network: PEEN. *Landsc. Ecol.*, **26**(3), 311–326.
- Jongman R.H.G., Bouwma I.M. & van Doorn A.M., 2006. The indicative map of the Pan-European Ecological Network in Western Europe : technical background report. *Alterra-rapport*, 1429.
- Jongman R.H.G., Külvik M. & Kristiansen I., 2004. European ecological networks and greenways. *Landsc. Urban Plan.*, **68**(2–3), 305–319.
- Laudelout A. & Paquet J.-Y., 2014. Les changements climatiques et les oiseaux : synthèse et impacts sur l’avifaune wallonne. *Aves*, **51**(4), 193–215.
- Lawson C.R. et al., 2014. Prevalence, thresholds and the performance of presence-absence models. *Methods Ecol. Evol.*, **5**(1), 54–64.
- Levey D.J. et al., 2005. Ecology: Effects of landscape corridors on seed dispersal by birds. *Science.*, **309**(80), 146–148.
- Levins R., 1969. Some Demographic and Genetic Consequences of Environmental Heterogeneity for Biological Control’ Some Demographic and Genetic Consequences of Environmental Heterogeneity for Biological Control’ **15**(3), 237–240.
- Lightbown, Patsy M. & Spada N., 2013. Corridors écologiques et conservation de la biodiversité, intérêts et limites pour la mise en place de la trame verte et bleue. *Sciences Eaux Territ.*, **3**, 34–39.
- Liu C., White M. & Newell G., 2013. Selecting thresholds for the prediction of species occurrence with presence-only data. *J. Biogeogr.*, **40**(4), 778–789.
- Lobo J.M., Jiménez-valverde A. & Real R., 2008. AUC: A misleading measure of the performance of predictive distribution models. *Glob. Ecol. Biogeogr.*, **17**(2), 145–151.
- Locquet A. & Clauzel C., 2018. Identification et caractérisation de la trame verte et bleue du PNR des Ardennes : comparaison des approches par habitat et par perméabilité des milieux. Identification and characterization of the Green and Blue infrastructures of the Regional Natural Park . *Cybergeo*, 0–27.
- MacArthur R.H. & Wilson E.O., 1967. *The Theory of Island Biogeography*, Princeton University Press, Princeton, New Jersey, 203.
- Mander Û., Külvik M. & Jongman R., 2003. Scaling in territorial ecological network. *Landschap* **20**(2), 113–127.
- Mateo R.G. et al., 2010. Profile or group discriminative techniques? Generating reliable species distribution models using pseudo-absences and target-group absences from natural history collections. *Divers. Distrib.*, **16**(1), 84–94.
- Melin E., 1997. La problématique du réseau écologique : Bases théoriques et perspectives d’une stratégie écologique d’occupation et de gestion de l’espace. *Colloq. Int., Le Réseau écologique*.

- Menard S., 2002. *Applied Logistic Regression Analysis, Second Edition*. Laughton, C.D. ed., Sage Publications, inc, Thousand oaks.
- Merow C., Smith M.J. & Silander J.A., 2013. A practical guide to MaxEnt for modeling species' distributions: What it does, and why inputs and settings matter. *Ecography (Cop.)*, **36**(10), 1058–1069.
- Muscarella R. et al., 2014. ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods Ecol. Evol.*, **5**(11), 1198–1205.
- Nations Unies, 1992. *Convention sur la diversité biologique*.
- Observation.be, 2020. Observation.be: Mission. <https://observations.be/pages/mission/>, (16/05/2020).
- Opdam P., Steingröver E. & Rooij S. Van, 2006. Ecological networks: A spatial concept for multi-actor planning of sustainable landscapes. *Landsc. Urban Plan.*, **75**(3–4), 322–332.
- Osborne P.E. & Suarez-Seoane S., 2002. Should data be partitioned spatially before building large-scale distribution models? **157**, 249–259.
- Palaoro A. V. et al., 2013. Niche conservatism and the potential for the crayfish *Procambarus clarkii* to invade South America. *Freshw. Biol.* **58**(7), 1379–1391.
- Paquet J. et al., 2013. La banque de données ornithoLogiques « courantes » d'aves : 50 ans de collecte d'observations de terrain. 9–19.
- Parlement Belge, 2019. Loi sur la Conservation de la nature, Bruxelles, 1–45.
- Parlement Européen, 2009. Directive 79/409/CE. *J. Off. l'Union Eur.*, 19.
- Pascual-Hortal L. & Saura S., 2007. Impact of spatial scale on the identification of critical habitat patches for the maintenance of landscape connectivity. *Landsc. Urban Plan.*, **83**(2–3), 176–186.
- Pearson R.G., Dawson T.P. & Liu C., 2004. Modelling species distributions in Britain: A hierarchical integration of climate and land-cover data. *Ecography (Cop.)*, **27**(3), 285–298.
- Pearson R.G. et al., 2007. Predicting species distributions from small numbers of occurrence records: A test case using cryptic geckos in Madagascar. *J. Biogeogr.*, **34**(1), 102–117.
- Pellissier V. et al., 2013. Assessing the Natura 2000 network with a common breeding birds survey. *Anim. Conserv.*, **16**(5), 566–574.
- Pepe M.S., 2000. Receiver operating characteristic methodology. *Stat. 21st Century*, **95**(449), 308–311.
- Peterson A.T., 2006. Uses and Requirements of Ecological Niche Models and Related Distributional Models. *Biodivers. Informatics*, **3**(0), 59–72.

- Peterson A.T. & Vieglais D.A., 2001. Predicting Species Invasions Using Ecological Niche Modeling: New Approaches from Bioinformatics Attack a Pressing Problem. *Bioscience*, **51**(5), 363.
- Philips S.J. & Dudik M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography (Cop.)*, **31**, 161–175.
- Phillips S.B. et al., 2006. Maximum entropy modeling of species geographic distributions. *Int. J. Glob. Environ. Issues*, **6**(2–3), 231–252.
- Phillips S.J. et al., 2009. Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecol. Appl.*, **19**(1), 181–197.
- Pontius R.G.J. & C.Schneider L., 2001. Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. *Agric. Ecosyst. Environ.*, **85**(1–3), 239–248.
- Powers D.M.W., 2011. Evaluation: from precision, recall and f-measure to roc, informedness, markedness & correlation. *Jour. of Machine Learning Tech.*, **2**(1), 37–63.
- Radoux J. et al., 2019. Improving ecotope segmentation by combining topographic and spectral data. *Remote Sens.* **11**(3), 1–15.
- Rich A.C., Dobkin D.S. & Niles L.J., 1999. Defining forest fragmentation by corridor width: the influence of narrow forest-dividing corridors on forest-nesting birds in southern New Jersey. *NCASI Tech. Bull.*, **2**(781), 440.
- Roberge J.M. & Angelstam P., 2004. Usefulness of the Umbrella Species Concept as a Conservation Tool. *Conserv. Biol.*, **18**(1), 76–85.
- Rosset S., 2004. Model Slection via the AUC. In: Brodley, C. ed. *International Conference Proceeding Series (ICPS)*. Association for Computing Machinery, New York.
- Santika T. & Hutchinson M.F., 2009. The effect of species response form on species distribution model prediction and inference. *Ecol. Modell.* **220**(19), 2365–2379.
- Saunders D.A. & Hobbs R.J., 1991. The role of corridors in conservation: what do we know and where do we go? *Nat. Conserv. 2 role corridors* (May 1991), 421–427.
- Saura S. & Pascual-Hortal L., 2007. A new habitat availability index to integrate connectivity in landscape conservation planning: Comparison with existing indices and application to a case study. *Landsc. Urban Plan.* **83**, 91–103.
- Saura S. & Rubio L., 2010. A common currency for the different ways in which patches and links can contribute to habitat availability and connectivity in the landscape. *Ecography (Cop.)*. **33**(3), 523–537.
- Sawyer S.C., Epps C.W. & Brashares J.S., 2011. Placing linkages among fragmented habitats: Do least-cost models reflect how animals use landscapes? *J. Appl. Ecol.* **48**(3), 668–678.
- Sieving K.E., Willson M.F. & De Santo T.L., 2000. Defining corridor functions for endemic birds in fragmented south-temperate rainforest. *Conserv. Biol.* **14**(4), 1120–1132.

- Smedshaug C.A. et al., 2002. The importance of the farmland-forest edge for area use of breeding hooded crows as revealed by radio telemetry. *Ornis Fenn.* **79**(1), 1–13.
- Smolik M.G. et al., 2010. Integrating species distribution models and interacting particle systems to predict the spread of an invasive alien plant. *J. Biogeogr.* **37**(3), 411–422.
- Spackman K.A., 1989. *Signal Detection Theory: Valuable Tools for Evaluating Inductive Learning*, Proceedings of the Sixth International Workshop on Machine Learning, Morgan Kaufmann Publishers, Inc., 160–163.
- Stockwell D.R.B. & Peterson A.T., 2002. Effects of sample size on accuracy of species distribution models. *Ecol. Modell.* **148**(1), 1–13.
- Story M. & Congalton R.G., 1986. Remote Sensing Brief Accuracy Assessment: A User's Perspective. *Photogramm. Eng. Remote Sensing* **52**(3), 397–399.
- Tewksbury J.J. et al., 2002. Corridors affect plants, animals, and their interactions in fragmented landscapes. *Proc. Natl. Acad. Sci. U. S. A.* **99**(20), 12923–12926.
- Urban d. & Keitt t., 2009. Landscape connectivity: a graph-theoretic perspective. *Ecology*, **82**(5), 1205–1218.
- Vanderwal J. et al., 2008. Selecting pseudo-absence data for presence-only distribution modeling : How far should you stray from what you know ? *Ecological Modelling*, **220**, 589–594.
- Vaughan I.P. & Ormerod S.J., 2005. The continuing challenges of testing species distribution models. *J. Appl. Ecol.*, **42**(4), 720–730.
- Veloz S.D., 2009. Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *J. Biogeogr.*, **36**(12), 2290–2299.
- Warren D.L. & Seifert S.N., 2011. Ecological niche modeling in Maxent: The importance of model complexity and the performance of model selection criteria. *Ecol. Appl.*, **21**(2), 335–342.
- Warren D.L. et al., 2014. Incorporating model complexity and spatial sampling bias into ecological niche models of climate change risks faced by 90 California vertebrate species of concern. *Divers. Distrib.*, **20**(3), 334–343.
- Wernham C.V. et al., 2002. *The Migration Atlas: movements of the birds of Britain and Ireland*, T. & A.D. Poyser.
- Wiens J.A., 1989. *The Ecology of Bird Communities: Foundations and Patterns*, Cambridge University Press, Cambridge, UK, 539.
- Yackulic C.B., et al., 2013. Presence-only modelling using MAXENT: When can we trust the inferences? *Methods Ecol. Evol.*, **4**(3), 236–243.
- Yates C.J. et al., 2010. Assessing the impacts of climate change and land transformation on *Banksia* in the South West Australian Floristic Region. *Divers. Distrib.*, **16**(1), 187–201.

- Zaniewski A.E., Lehmann A. & Overton J.M.C., 2002. Predicting species spatial distributions using presence-only data: A case study of native New Zealand ferns. *Ecol. Modell.* **157**(2–3), 261–280.
- Zeng Y., Low B.W. & Yeo D.C.J., 2016. Novel methods to select environmental variables in MaxEnt: A case study using invasive crayfish. *Ecol. Modell.* **341**, 5–13.