
Prédiction des zones favorables à l'installation de l'espèce exotique envahissante *Ambrosia artemisiifolia* L. en région wallonne

Auteur : Delforge, Adrien

Promoteur(s) : Monty, Arnaud

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/2268.2/10622>

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.

**PREDICTION DES ZONES FAVORABLES A
L'INSTALLATION DE L'ESPECE EXOTIQUE
ENVAHISSANTE AMBROSIA ARTEMISIIFOLIA L. EN
REGION WALLONNE**

ADRIEN DELFORGE

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

ANNÉE ACADÉMIQUE 2019-2020

PROMOTEUR: A. MONTY

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

« Le présent document n'engage que son auteur. »

**PREDICTION DES ZONES FAVORABLES A
L'INSTALLATION DE L'ESPECE EXOTIQUE
ENVAHISSANTE AMBROSIA ARTEMISIIFOLIA L. EN
REGION WALLONNE**

ADRIEN DELFORGE

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

ANNÉE ACADÉMIQUE 2019-2020

PROMOTEUR: A. MONTY

Remerciements

Nous tenons à remercier l'Observatoire Wallon des Ambroisies, sans qui ce travail n'aurait pas vu le jour. Toute notre reconnaissance à notre promoteur Monsieur A. Monty pour ses nombreuses relectures et ses précieux conseils, ainsi qu'à A. Bourdouxhe et T. Copée pour leur expertise en matière de modélisation.

Enfin, je remercie les personnes ayant aidé à la correction du document à savoir Danielle Delattre, Jean-François Delforge et Dasha De Ruyver.

Résumé

L'ambrosie à feuilles d'armoise (*Ambrosia artemisiifolia* L.) est une espèce exotique envahissante originaire d'Amérique du Nord qui a conquis l'Europe depuis le début du 20^{ème} siècle. Malgré diverses observations en Belgique, l'espèce est peu présente et peu connue en région wallonne. L'espèce cause de nombreux dégâts d'ordre sanitaire et économique, étant donné qu'elle est une adventice de culture très allergénique. La modélisation spatiale des zones favorables à l'installation permettrait une meilleure sensibilisation, ainsi qu'un ciblage des échantillonnages et des luttes contre l'espèce.

La présente étude regroupe différentes méthodes de modélisation utilisées dans la littérature afin de mieux comprendre les difficultés rencontrées par une modélisation des espèces invasives. Par utilisation des points d'occurrence et des rasters bioclimatiques Worldclim de la zone d'indigénat, une première modélisation a été réalisée par extrapolation sur le territoire wallon. Une extrapolation temporelle a aussi été réalisée sur le modèle, afin de constater l'impact du changement climatique sur la distribution potentielle de l'ambrosie à feuille d'armoise. Une seconde modélisation a été effectuée à partir des données environnementales de 'Lifewatch Ecotope' et les données de présence wallonnes, malgré la faible robustesse de ces dernières. La validation du modèle s'est faite avec l'indice AUC, et les résultats des deux modèles ont montré des zones favorables sur une grande partie du territoire wallon en province du Hainaut, de Namur, du Brabant Wallon et de Liège.

Les variables contraignant le plus le modèle extrapolé sont les températures et les précipitations estivales ainsi que la variation des précipitations. Les variables contraignant le plus le modèle indépendant sont les températures minimales hivernales, la distance aux chemins de fer et la présence de sols à drainage moyen. L'extrapolation temporelle a démontré que la quasi-totalité du territoire wallon sera favorable à la colonisation de l'ambrosie à feuilles d'armoise, à l'exception de l'est de la province du Luxembourg et l'est de la province de Liège. Le modèle indépendant a été moins performant que le modèle extrapolé, essentiellement à cause de la faiblesse du jeu de données utilisé.

Il est important de rester attentif à la prolifération de l'ambrosie à feuilles d'armoise en région wallonne. Si elle n'est pas encore très nocive dans la région, les dégâts qu'elle peut engendrer suffisent pour sensibiliser au maximum les acteurs liés à l'introduction de l'espèce et à sa gestion. Il est essentiel de multiplier les points d'observations en région wallonne afin de mieux comprendre la distribution de l'espèce et lutter au mieux contre elle dans le futur.

Abstract

The common ragweed (*Ambrosia artemisiifolia* L.) is an invasive alien species native to North America that has conquered Europe since the turn of the 20th century. Despite various observations in Belgium, the species has few presence data and is not very well-known in the Walloon region. The species causes extensive health and economic damage, as it is a highly allergenic crop weed. Spatial modelling of areas favourable to its installation would allow better sensitization, as well as targeted sampling and management of the species.

This study brings together different modeling methods used in the literature in order to better understand the issues encountered by modeling invasive species. Using the presence points and the Worldclim bioclimatic rasters of the native area, a first modeling was carried out by extrapolation over the Walloon territory. A temporal extrapolation was also performed on the model, to analyse the impact of climate change on the potential distribution of common ragweed. A second modeling was carried out using environmental data from 'Lifewatch Ecotope' and the Walloon presence data, despite their poor robustness. The model was validated with the AUC index, and the results of the two models showed favorable areas over a large part of the Walloon territory in the provinces of Hainaut, Namur, Brabant Wallon and Liège.

The variables constraining the extrapolated model the most are summer temperatures and precipitation as well as the variation in precipitation. The variables constraining the independent model the most are the minimum winter temperatures, the distance to railways and the presence of soils with average drainage. The temporal extrapolation has shown that almost all of the Walloon territory will be favorable to the colonization of the common ragweed, with the exception of the east of the province of Luxembourg and the east of the province of Liège. The independent model performed less well than the extrapolated model, mainly because of the weakness of the data set used.

It is important to remain attentive to the proliferation of common ragweed in the Walloon region. While it is not yet very harmful in the region, the potential damage it can cause is enough to raise as much awareness as possible among those involved in the introduction of the species and its management. It is essential to increase the number of observation points in the Walloon region in order to better understand the distribution of the species and to better fight against it in the future.

Table des matières

Introduction.....	1
Les espèces exotiques envahissantes (EEE)	1
La modélisation spatiale et la distribution des espèces.....	3
La modélisation des espèces exotiques envahissantes : différences et challenges	4
Revue bibliographique	7
Sélection d'un modèle.....	7
Revue des différentes méthodes d'optimisation d'un modèle de niche écologique (ENM) pour une espèce exotique envahissante (EEE)	9
Choix d'un algorithme	11
Réduction des biais d'échantillonnage.....	12
Génération des pseudo-absences	13
Validation du modèle	17
Matériel et méthode	21
L'espèce étudiée : <i>Ambrosia artemisiifolia</i> L.....	21
Biologie de l'espèce.....	21
Répartition de l'espèce dans le monde.....	23
Prédictions pour sa future répartition	25
Construction d'un modèle extrapolé	26
Données utilisées	26
Méthode.....	28
Construction d'un modèle indépendant	30
Données utilisées	30
Méthode.....	30
Résultats.....	33
Modèle extrapolé	33
Modèle indépendant.....	40
Discussion.....	45
Le modèle extrapolé.....	45
Le modèle indépendant.....	48
Comparaison des cartes et futur de l'ambrosie en Wallonie	49
Manipulations compromises par la situation sanitaire.....	51
Conclusion	53
Bibliographie.....	54
Annexe.....	64

Introduction

Les espèces exotiques envahissantes (EEE)

Les espèces exotiques envahissantes (EEE) sont considérées comme l'un des plus grands challenges écologiques de notre époque. Une EEE se définit comme une espèce naturalisée, présente en grand nombre et se développant sur une aire considérable, causant des effets négatifs. Ces invasions sont dès lors considérées comme l'une des principales menaces pour l'environnement et la biodiversité. Tous les taxons sont concernés par ce problème, dont les végétaux.

Les invasions biologiques ne sont pas des phénomènes récents, mais l'action de l'homme en a fortement augmenté la fréquence (Richardson et al., 2000), permettant ainsi à des espèces de coloniser des territoires qui leur étaient jusque-là impossible à atteindre. De cette manière, les échanges internationaux ont favorisé l'expansion des plantes dites invasives, ainsi que d'autres activités humaines telles que l'agriculture ou les loisirs, et n'ont épargné aucune région dans le monde. Dans un futur proche, le changement climatique devrait redistribuer les aires de répartition de toutes les espèces, dont ces espèces exotiques envahissantes, ce qui constitue un nouveau challenge dans l'expansion des EEE.

Ces nouvelles introductions, qu'elles soient volontaires ou accidentelles, ont permis à certaines espèces de s'installer de se reproduire dans des proportions considérables, causant alors de nombreux problèmes. Premièrement, l'arrivée de ces espèces impliquent un changement dans l'écosystème, ce qui peut perturber son équilibre. En résultent alors plusieurs conséquences écologiques, qui peuvent aller de la perte de la biodiversité à l'interruption des stades de succession naturels. De plus, ces espèces peuvent avoir d'autres effets néfastes, qu'ils soient économiques (adventices de cultures), sociaux (perte d'identité culturelle) ou sanitaires (allergies, brûlures). Se débarrasser de ces EEE est donc devenu une priorité, mais demande bien souvent des efforts considérables qu'ils soient humains ou financiers, pour réduire ou maintenir une population.

Pour devenir une EEE, une espèce doit parvenir à surmonter plusieurs barrières environnementales (Williamson, 1996). En effet, la majorité des espèces disparaissent après leur introduction, face à des conditions abiotiques et biotiques défavorables. Une fois introduite en dehors de sa niche d'origine, l'espèce doit d'une part survivre, mais aussi pouvoir se reproduire et se disperser efficacement. Ce n'est que lorsque ces conditions sont remplies que l'espèce devient naturalisée (Richardson et al., 2000), et peut alors devenir une espèce exotique envahissante, si elle se disperse rapidement et dans des proportions telles qu'elle devient un problème. Elle colonise alors de nouveaux habitats (Alexander and Edwards, 2010; Monty and Mahy, 2009) mais elle est rapidement confrontée à ses limites environnementales, et retrouve une situation plus équilibrée. On dit alors que l'espèce n'est pas à l'équilibre dans son environnement lors des premières phases d'invasion. Identifier ces limites d'invasion est donc primordial pour comprendre la distribution d'une EEE. Il peut arriver que malgré une naturalisation ratée, une espèce prospère dans une région défavorable. Cette présence est due à une introduction répétée, par exemple à cause de graines contaminées en milieu agricole. On parle alors de '*casual species*' (Pyšek et al., 2004).

On distingue d'ordinaire trois phases de colonisation d'une EEE : la phase de latence lors de laquelle l'espèce se disperse relativement lentement et ne pose pas de problème majeur ; la phase exponentielle ou de colonisation où l'espèce devient problématique de par sa dispersion accélérée ; et la phase de limite de capacité, où l'espèce a colonisé tous les milieux qui lui étaient favorables. Il est

donc pertinent de lutter contre une espèce le plus tôt possible dans sa courbe de colonisation, afin de réduire au maximum les coûts et les dégâts occasionnés par l'espèce.

L'un des facteurs les plus importants à analyser lorsqu'on étudie les EEE est l'étendue de sa distribution, ou du moins l'étendue des habitats sensibles à son invasion. En effet, identifier les zones sensibles à une EEE et les facteurs qui les définissent permet de mieux comprendre l'espèce, et favoriser donc la lutte contre celle-ci, en ciblant les zones à échantillonner ou en investissant dans ces zones sensibles.

Afin de cibler correctement ces zones, il est important de comprendre ce que l'on désire identifier. Il est dès lors nécessaire de comprendre quel type de niche écologique est ciblée (Elith, 2017), en se référant aux définitions de Hutchinson (Hutchinson, 1957). Ainsi, la niche fondamentale est la niche inhérente à l'espèce, limitée par tous les facteurs biotiques et abiotiques qui pourraient affecter négativement le fitness de l'espèce, tels que la lumière, les variables édaphiques ou climatiques, tandis que la niche réalisée est la niche effective, celle que l'espèce utilise à cause des contraintes qu'elle rencontre, comme l'activité humaine ou les prédateurs (Colwell and Rangel, 2009). D'autres auteurs (Pulliam, 2000) ont affiné la définition en y ajoutant la théorie '*source-sink*', et les limites de dispersions. De cette manière, il est tout à fait possible que la niche réalisée soit moins contraignante que la niche fondamentale, notamment en début d'invasion dans le cas des EEE, où l'espèce n'est pas en équilibre dans son environnement.

Les zones sensibles à identifier sont donc liées à la niche réalisée. Le lien entre la niche écologique et la distribution de l'espèce n'est plus à démontrer, la distribution d'une espèce étant l'expression de sa niche écologique dans l'espace (Sexton et al., 2009). Cependant, la dynamique des espèces crée une dissonance entre la niche et la réelle distribution des espèces (Schurr et al., 2012). En effet, les espèces se trouvent parfois dans des environnements non favorables, et un changement environnemental causera des situations de non équilibre, créant des populations absentes des zones favorables ou présentes dans des zones qui ne le sont plus (extinction future) (Schurr et al., 2012).

Enfin, une complication supplémentaire est le phénomène de '*range-shift*' ou '*niche shift*' que peut présenter une espèce exotique envahissante. En effet, des nouvelles conditions peuvent provoquer une adaptation de la niche écologique (Colwell and Rangel, 2009), par des évolutions génétiques liées au faible nombre d'individus importés par exemple.

Une des manières les plus largement répandues pour identifier les zones favorables à une espèce est l'utilisation de modèles de distribution d'espèce, appelés couramment '*Species Distribution Model*' (SDM) ou '*Ecological Niche Model*' (ENM).

La modélisation spatiale et la distribution des espèces

La modélisation spatiale

L'intérêt pour la description, la compréhension et la prédiction géographique et environnementale de la distribution d'une espèce est très ancienne (Grinnell, 1917). Depuis la fin des années 1990, le nombre d'études utilisant la corrélation entre les données environnementales et d'occurrence a augmenté considérablement (Thuiller et al., 2009). En effet, des centaines d'articles sortent chaque année concernant ou utilisant ces outils (Lobo et al., 2010). Il a été prouvé qu'il existe une forte relation entre l'environnement et la distribution des espèces (Araújo and Guisan, 2006; Pulliam, 2000), ce qui donne du poids aux modélisations basées sur des variables environnementales. Le nombre de données disponibles en ligne est très certainement la raison de cette augmentation, que ce soit pour les données d'occurrence des espèces, via notamment le site GBIF (Edwards, 2004) ou pour les données bioclimatiques, via la base de données Worldclim (Fick and Hijmans, 2017) par exemple, très largement utilisée en modélisation spatiale.

Plusieurs types de modèles existent pour modéliser la distribution d'une espèce ou les zones favorables à sa dispersion. Il est possible de catégoriser ces modèles en deux grands groupes : les modèles mécanistes et les modèles de corrélation. Les modèles mécanistes vont essentiellement se baser sur la biologie et l'écologie, les traits fonctionnels de l'espèce étudiée (Buckley et al., 2010), permettant alors de se rapprocher de la niche fondamentale de l'espèce. Ces modèles dépendent plus des mécanismes et traits écologiques, plutôt que de motifs particuliers, demandant alors des connaissances approfondies sur l'espèce, et de nombreuses données d'entrée pour le modèle.

A l'inverse, les modèles de corrélation vont totalement ignorer l'écologie de l'espèce, pour se concentrer exclusivement sur la comparaison entre les absences et les présences, sur plusieurs variables environnementales données. Bien que les modèles mécanistes sont généralement conseillés pour l'identification des zones envahissables ou de la distribution d'une EEE (Buckley et al., 2010; Gallien et al., 2010), peu de méthodologies précises existent. En effet, les modèles mécanistes sont construits au moyen de programmes spécifiques ou selon des méthodologies propres aux études qui les utilisent, mais aucune méthode ne s'est démarquée au fil des études réalisées. Les deux modèles peuvent bien entendu se mêler, et la sélection des variables en entrée peut par exemple être un moyen d'intégrer des connaissances de la biologie de l'espèce dans un modèle de corrélation.

Plusieurs auteurs ont comparé les deux types de modèles (Dormann et al., 2012; Hijmans and Graham, 2006; Morin and Thuiller, 2009). Bien que les deux modèles semblent donner des résultats similaires pour une région de référence, il existe de nombreuses différences lorsqu'on projette le modèle dans un nouvel espace ou un autre temps.

Les modèles de distribution d'espèce

Concernant les modèles de corrélation, il s'agit des modèles de distribution d'espèce ou '*Species Distribution Models*' (SDM ; (Elith and Leathwick, 2009; Guisan and Zimmermann, 2000)) ou de niches écologiques (ENM , (Harrison, 1997; Peterson, 2006a)), très largement utilisés dans la littérature. Avant d'aller plus loin, il semble important de s'attarder sur la terminologie de cet outil. Les deux outils peuvent être définis de la manière suivante : « *un modèle ou processus qui a pour but de produire une carte des zones qui sont potentiellement favorables pour une espèce* » (Elith, 2017). Le but des deux outils est toutefois différent : les SDMs auront pour but de cartographier la distribution de l'espèce, quand les ENMs cartographieront sa niche écologique. Il serait alors plus juste de considérer les ENMs comme un intermédiaire nécessaire pour atteindre un SDM. Selon le but de l'étude, la terminologie

doit donc changer. Même si les deux sont issus des mêmes calculs, il est important de les distinguer pour une interprétation exacte (Peterson and Soberón, 2012).

Plusieurs études (Peterson and Soberón, 2012) insistent sur la distinction entre les SDM et les ENM. En effet, beaucoup d'études confondent les deux termes, et assimilent les SDM aux ENM. Plusieurs critiques générales sont adressées à ces modèles, notamment sur la réelle relation entre le résultat obtenu et la niche écologique (Soberón, 2010), ou le manque de données écologiques dans ces processus (Austin, 2002). Pour éviter la confusion, certains auteurs utilisent le terme '*Bioclimatic envelope models*' (Araujo and New, 2007; Araújo and Peterson, 2012; Bedia et al., 2013; Jeschke and Strayer, 2008), qui lui-même est aussi controversé, les variables les plus contraignantes n'étant pas toujours d'ordre bioclimatique.

Enfin, il peut être utile de rappeler que les SDMs et ENMs ne prennent pas en compte les situations de non-équilibre et l'échelle locale, deux facteurs très importants dans la distribution de l'espèce. Ces modèles produisent donc des niches biaisées, notamment sur les projections futures.

La modélisation des espèces exotiques envahissantes : différences et challenges

La modélisation dans la lutte contre les EEE

Le contrôle et la gestion des EEE sont souvent accompagnés d'analyse de risques, qui permettent de mieux comprendre l'écologie de l'espèce. Les modélisations spatiales sont alors un outil majeur dans ces analyses (Jeschke and Strayer, 2008; Jiménez-Valverde et al., 2011; Václavík and Meentemeyer, 2009), en construisant une carte des zones favorables à sa naturalisation, ou de sa potentielle distribution (Jiménez-Valverde et al., 2011). Les avantages de ce type de carte sont nombreux et logiques : il est possible de comprendre les zones à risques pour une EEE donnée, pour ainsi cibler les zones à échantillonner ou diriger la lutte contre l'espèce (Brunel et al., 2010).

Cet avantage dans la lutte contre les EEE a été bien comprise ; ainsi plusieurs organisations ou gouvernements investissent dans ce genre d'outils. En Europe, on peut par exemple citer '*Pratique*', une organisation spécialisée dans l'analyse de risques des EEE, qui produit des cartes spécialisées ou l'autorité européenne de sécurité des aliments (EFSA).

Plusieurs études se sont penchées sur la réponse des espèces exotiques envahissantes au changement climatique (Ricciardi et al., 2017; Ziska and Caulfield, 2000). Pour ce faire, les ENMs sont des outils qui sont largement utilisés pour prédire les zones favorables à l'établissement d'une espèce dans le temps ou l'espace (Shabani et al., 2016). Dans le cadre d'EEE, le modèle est calibré avec la zone d'indigénat ou une zone de référence, puis projeté dans la zone envahie ou avec des variables associées au temps futur considéré. Cette méthode est régulièrement critiquée, les modèles étant parfois peu transférables, ou à cause du phénomène de 'niche shift' qui touche certaines espèces exotiques envahissantes.

Il est important de noter que les modèles corrélatifs n'ont pas été construits dans le but de modéliser une distribution d'EEE ou des zones sensibles à une espèce. Ces modèles ont été conçus pour utiliser les données d'une étendue définie, avec une espèce en équilibre dans son environnement (Elith and Leathwick, 2009; Phillips et al., 2006). Utiliser des modèles de corrélation pour une étude traitant des espèces exotiques envahissantes demande donc un effort dans l'élaboration du modèle, mais aussi dans l'interprétation.

Plusieurs revues et tutoriels ont été réalisés depuis le début des années 2000 (Austin, 2002; Elith and Leathwick, 2009; Franklin, 2010; Guisan and Thuiller, 2005; Guisan and Zimmermann, 2000; Schröder, 2008). Plusieurs modèles de corrélation peuvent être comparés, afin de choisir celui qui sied le mieux aux objectifs de l'étude. Dans le cadre d'un modèle choisi pour identifier la distribution d'une EEE ou les zones sensibles à une invasion, certains auteurs utilisent le terme iSDM (Brummer et al., 2013; Gallien et al., 2012; Václavík and Meentemeyer, 2009). Le but de ces études est soit d'identifier les zones sensibles à l'EEE, soit de déterminer les variables qui conduisent la dispersion. Malgré le fait que ces espèces ne soient pas en équilibre dans leur environnement, les modèles spatiaux restent un moyen rapide et efficace pour apporter une information supplémentaire dans la lutte contre les EEE.

La modélisation des EEE : points d'attention

Si la modélisation de distribution d'espèce ou de niches écologiques reste un outil majeur et largement répandu dans les études de conservation (notamment la gestion des espèces rares et des EEE), il existe une multitude de paramètres qui peuvent être mal utilisés ou interprétés par les utilisateurs (Elith, 2017; Halvorsen et al., 2016). En effet, les procédures de modélisation (tel Maxent) devenant de plus en plus '*user-friendly*', une grande majorité des modélisations spatiales ne porte pas assez attention à certaines limites de l'outil de modélisation. Outre la situation de déséquilibre dans l'environnement, il existe en effet plusieurs sujets d'attention concernant la modélisation spatiale, propres au phénomène des espèces exotiques envahissantes.

Premièrement, même si le problème ne concerne pas que les EEE, les données d'occurrence sont considérées complètes et précises ou du moins suffisamment représentatives de la distribution de l'espèce par le modèle (Phillips et al., 2006). Cette affirmation de départ est très souvent fautive pour plusieurs raisons. D'une part, l'origine des données de présence peut être multiple : données historiques via des planches de musées, échantillonnage ciblé de l'espèce ou encore utilisation de jeux de données de type Gbif, qui utilisent des données encodées par des particuliers. Selon l'origine des données d'occurrence, il faut donc anticiper les imprécisions géographiques, mais aussi taxonomiques (Elith and Leathwick, 2009). De plus, la date des données est une information importante, les données d'un passé trop lointain n'étant potentiellement plus d'actualité.

Il est d'ordinaire admis d'utiliser les données de la zone d'indigénat pour modéliser des zones sensibles d'une région envahie par une EEE, ce qui minimise cette problématique. En effet, plus l'espèce est présente depuis longtemps dans une région, plus on considère l'espèce comme correctement échantillonnée, spécialement lorsqu'il s'agit d'une EEE, ces dernières étant très souvent généralistes et donc plus facilement observables.

Un autre gros problème lié à l'échantillonnage est qu'il n'y a souvent aucune information sur l'effort d'échantillonnage (Barbet-Massin et al., 2010; Jarnevich et al., 2015; Phillips et al., 2009) et plusieurs biais spatiaux peuvent apparaître dans les données d'occurrence, biaisant ainsi le modèle. Un travail de filtrage spatial ou de modification des données d'entrée est donc souvent nécessaire.

La collecte des données environnementales est bien souvent un autre problème pour les chercheurs. Il existe une multitude de données, notamment bioclimatiques tels Worldclim (Fick and Hijmans, 2017) qui propose une résolution très fine (~500m) mais le défaut de ces variables est qu'elles présentent une forte corrélation entre elles. La majorité des modèles utilisent des variables bioclimatiques, en ignorant les autres variables environnementales tels que la couverture de sol ou la pollution lumineuse parfois essentielle dans la distribution de l'espèce.

La sélection des variables reste un obstacle majeur. En effet, il existe une multitude de méthode pour les sélectionner (Warren et al., 2014), selon la précision ciblée du modèle, ou son utilisation. Certains

vont préférer sélectionner directement les variables en fonction de l'écologie de l'espèce (Kearney et al., 2008; Rödder et al., 2009), quand d'autres sélectionneront plutôt selon la participation au modèle. Il est important de mettre de côté les variables auto-corrélées et il est donc essentiel de sélectionner un set de variables peu corrélées, avec un seuil de corrélation adapté à l'étude. De plus, une bonne connaissance de l'espèce est primordiale, afin de conserver les variables les plus contraignantes pour la distribution (Jarnevich et al., 2015).

Il existe encore une multitude d'autres problèmes liés à la modélisation de la distribution potentielle d'une EEE, ou des zones sensibles à l'espèce. Parmi eux, on retrouve la transférabilité du modèle. Ce concept décrit la capacité d'un modèle à être extrapolé dans une région différente ou dans une différente temporalité, et est donc au cœur des modélisations concernant les EEE. Cette extrapolation des modèles est souvent remise en cause par les auteurs (Jarnevich et al., 2015), les modèles étant construits pour un moment donné dans un espace défini. Certains types d'algorithme sont donc conseillés pour une meilleure transférabilité, ainsi que la réduction du nombre de variables en entrée, bien que le domaine soit encore peu exploré. Les modèles construits avec un nombre réduit de variables sont aussi à préférer pour améliorer la transférabilité (Radosavljevic and Anderson, 2014; Warren and Seifert, 2011), ainsi que des variables connues pour contraindre effectivement la niche de l'espèce.

Enfin, d'autres paramètres tels que la validation du modèle ou le choix de l'algorithme sont au cœur de l'optimisation d'un modèle permettant d'identifier des zones sensibles à une EEE. De cette manière, il est pertinent de réaliser une revue bibliographique des différentes méthodes utilisées dans la littérature pour modéliser des zones sensibles à une EEE. Dans cette optique, la première partie de cette étude s'attardera sur les différents choix à effectuer avant et pendant la modélisation. Ensuite, le deuxième objectif sera de réaliser une modélisation sur un sujet d'étude : *Ambrosia artemisiifolia* L. ou ambrosie à feuilles d'armoise (Figure 1).

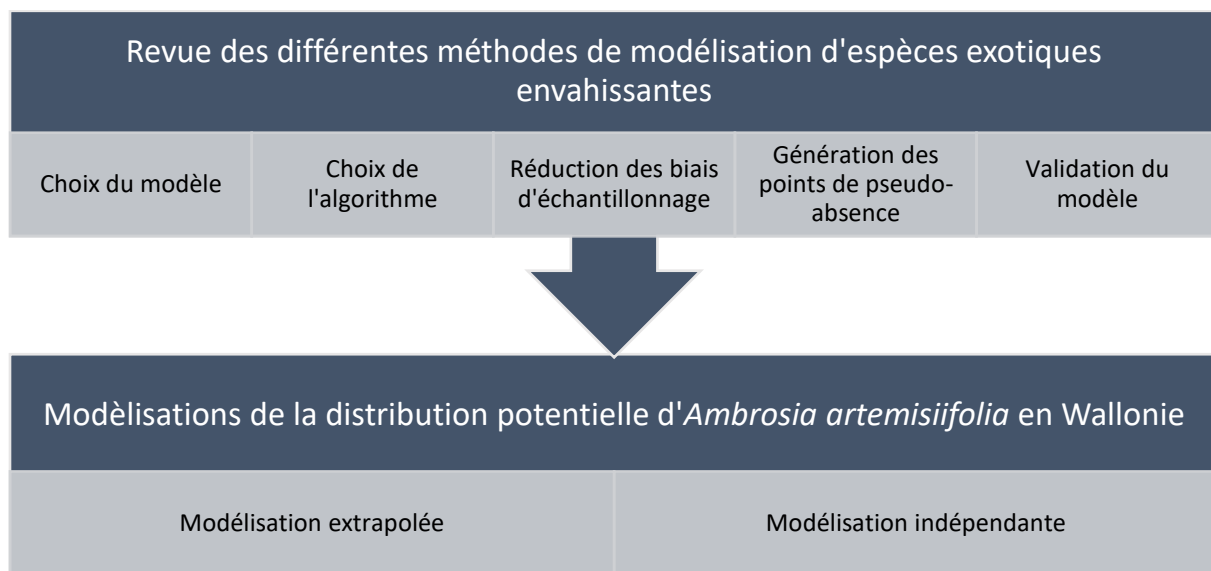


Figure 1- Schéma des différents objectifs poursuivis par l'étude

Revue bibliographique

Afin de réaliser le modèle le plus adapté possible, il est impératif de comprendre les différentes options possibles dans la construction du modèle. Ce secteur d'activité est constamment en étude et une multitude d'articles paraissent chaque année afin de comparer les meilleurs paramètres et méthodes dans des circonstances données. Cette revue bibliographique n'a pas pour but d'être exhaustive, mais de rassembler les principales options disponibles afin de faciliter le choix de méthode pour une modélisation spatiale. Un focus particulier sera mis sur la modélisation de la distribution potentielle des EEE.

Sélection d'un modèle

Afin d'identifier la distribution potentielle d'une espèce ou les zones favorables à sa dispersion dans le cadre d'une colonisation d'EEE, il existe plusieurs types de modèles. Comme mentionné précédemment, on distingue généralement les modèles corrélatifs et les modèles mécanistes.

Cependant, il existe une multitude de modèles différents pour étudier la distribution spatiale d'une espèce. Ainsi, parmi les modèles mécanistes, on compte les modèles qui vont identifier la distribution actuelle sur base des données d'occurrence et les modèles qui vont intégrer la dynamique de l'espèce, afin de concorder au mieux à la distribution réelle. En effet, la distribution d'une espèce est dynamique et varie donc en fonction des interactions abiotiques et biotiques ou des perturbations par exemple. De cette manière, les '*dynamics range models*' (DRMs), tels les modèles '*Rangeshifter*' de Bocedi et al. (2014) sont très utiles et plus fidèles à la réalité (Lustig et al., 2017; Sullivan et al., 2012). Ils étudient la réponse environnementale par rapport au taux démographique d'une espèce et permettent de projeter le modèle dans le temps. Ce type de modèle est particulièrement difficile à obtenir, de par les efforts statistiques considérables qu'il demande, les hypothèses de départ exigeantes (espèce en équilibre, données d'occurrence complètes) et les connaissances de l'espèce importantes (Schurr et al., 2012). De plus, ces modèles ne permettent que rarement plusieurs couches environnementales, ce qui est préjudiciable lorsqu'une espèce est sensible à plusieurs facteurs différents.

Les modèles dynamiques ne semblent donc pas convenir aux EEE, alors que la problématique de dispersion est essentielle dans leur étude. Certains modèles peuvent néanmoins s'adapter aux EEE, lorsque les données d'occurrence sont précises géographiquement et accompagnées d'une composante temporelle. Il est alors possible de modéliser la dynamique d'expansion de l'espèce, et de l'appliquer dans le futur. Enfin, il est aussi possible d'extrapoler la dynamique d'expansion d'une espèce d'un endroit connu à une zone d'étude, afin de prédire l'expansion dans une nouvelle région. Bien qu'utiles, ces modèles sont très gourmands en données d'entrée et donc rarement réalisables. De plus, l'expansion d'une espèce peut varier d'une région à l'autre, de par la variation comportementale de l'espèce, ou la différence de comportement anthropique, souvent très lié à l'expansion d'une EEE.

Ainsi, les modèles les plus utilisés en modélisation spatiale sont les ENM ou SDM, modèles corrélatifs qui permettent d'identifier, dans la cas d'une expansion d'EEE, les zones favorables à l'établissement de cette espèce.

Différentes manières de construire un modèle de type SDM

Bien que le choix du modèle soit effectué, il existe encore plusieurs manières de le construire. En effet, dans le cadre des EEE, il existe deux méthodes principales. D'une part, travailler de la même manière que pour les espèces en équilibre et indigènes, c'est-à-dire prendre en données d'entrée les données

d'occurrence de la zone d'étude. Le modèle est donc indépendant, toutes les données étant issues de la zone d'étude. Ce type de modèle est déconseillé pour les EEE, particulièrement lorsque l'espèce n'est pas présente depuis suffisamment longtemps (Elith, 2017). En effet, une espèce qui n'est pas à l'équilibre dans son environnement présente des caractéristiques spatiales très différentes, qui peuvent fausser les résultats. De plus, les données sont très incomplètes et l'échantillonnage généralement peu représentatif. Certaines études récentes tendent cependant à montrer que ces modèles peuvent prédire correctement la distribution potentielle d'une espèce, même lorsque l'espèce n'est pas en équilibre dans son environnement (Fand et al., 2020; Yan et al., 2020). Ce type de modèle permet une résolution généralement plus élevée, la zone d'étude étant plus restreinte. De plus, les données environnementales sont souvent plus facilement disponibles et plus précises.

De manière générale, il est donc conseillé de faire une extrapolation de modèle pour les EEE (Elith, 2017; Liang et al., 2018). Un modèle est donc créé à partir des informations de la zone native ou d'une zone où l'espèce est considérée en équilibre dans son environnement. Le modèle généré est ensuite appliqué à la zone d'étude. Les données d'entrée sont donc souvent plus fiables, mais le nombre de variables en entrée est néanmoins plus faible, dû à l'indisponibilité de certaines variables qui sont parfois essentielles à l'étude de la distribution de l'espèce. Seules les variables bioclimatiques sont utilisées, la majorité du temps, lorsque ce type de modèle est utilisé.

Les deux modèles semblent donc apporter leur lot d'avantages et d'inconvénients et il peut dès lors être intéressant de les interpréter tous les deux.

Revue des différentes méthodes d'optimisation d'un modèle de niche écologique (ENM) pour une espèce exotique envahissante (EEE)

La construction d'un ENM ou SDM peut se fractionner en trois étapes : le pré-processing, le processing et le post-processing (Figure 3) (Andrade et al., 2020). Afin de mieux appréhender les différentes options d'optimisation du modèle, il est pertinent de s'attarder sur ces étapes.

Lors de la phase de pré-processing, toutes les manipulations nécessaires au bon fonctionnement de l'algorithme principal sont effectuées. Parmi celles-ci, seuls l'obtention des données de présence, des rasters environnementaux ou climatiques et la récupération des données d'absence (à l'exception de quelques algorithmes) sont obligatoires. Selon l'algorithme utilisé, ces données d'absence peuvent être récoltées sur le terrain (on parle alors d'absences 'vraies') ou générées. Dans ce dernier cas, ces points sont appelés points de pseudo-absence, car générés artificiellement. Ces points de pseudo-absence, lorsqu'ils sont bien plus nombreux que les points de présence, constituent alors un background à ces données. L'ensemble des points de pseudo-absence devient alors des points de background (Figure 2). Dans le cas où les points de pseudo-absence sont en nombre comparable aux points de présence, leur appellation ne change pas. Cette génération peut faire l'objet de nombreuses manipulations, détaillées plus en aval dans le document.

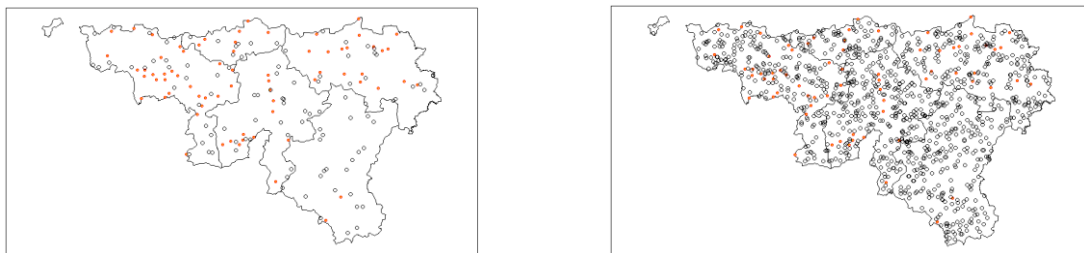


Figure 2 - Comparatif des deux principales options de nombre de génération de background. Les points de présence sont identifiés en rouge, tandis que les points de pseudo-absence sont en blanc bordés de noir. L'image de gauche présente un nombre de pseudo-absences similaire au nombre de points de présence, tandis que l'image de droite en présente un nombre dix fois supérieur. Ces points de pseudo-absence constituent dans ce dernier cas les points de background ou plus simplement le background.

Les données de présence, ainsi que les données environnementales peuvent être aussi manipulées. La résolution des cartes de prédictions sera définie par les rasters environnementaux, qui doivent être de même résolution, de même étendue, de même système de coordonnées et contenant des valeurs soit continues, soit catégorielles. Les pixels de ces cartes seront appelés 'cellule', constituant la plus petite unité géographique de l'étude. Ces rasters environnementaux peuvent être enfin l'objet de sélections, afin de ne garder que les variables les plus pertinentes. Enfin, c'est pendant cette phase que le partitionnement des données en vue de la validation, ou de l'ajout de données supplémentaires tels qu'un vecteur des zones accessibles sont effectuées.

Lorsque toutes les données sont importées, générées et manipulées comme souhaité, la phase de processing a lieu. L'algorithme choisi récupère les données d'entrée, et génère alors un modèle en comparant les valeurs de présence et d'absence avec les données environnementales ou climatiques. Ce modèle est ensuite utilisé pour créer une carte des valeurs prédictives, dont chaque cellule a une valeur de 0 à 1. Plus la valeur est élevée, plus l'espèce a de chance d'être présente. Cependant, ces valeurs prédictives ne constituent pas une échelle de probabilité. En effet, le calcul d'un seuil permettra de créer une carte binaire afin de distinguer les zones favorables ou non, mais une valeur

prédictive de 0.5 ne signifie pas une probabilité de 50% de présence. Lors de cette phase, différents outils d'analyse peuvent être aussi générés, tels que la contribution des variables au modèle, ou les réponses écologiques de celle-ci selon les valeurs prédictives.

Enfin, la phase de post-processing comprend les dernières manipulations suite à la création du modèle. La création d'une carte de présence/absence ou favorable/défavorable peut alors se faire, après le calcul d'un ou plusieurs seuils. La validation du modèle au moyen de différents indices métriques peut aussi se faire dans cette phase, ainsi que l'extrapolation du modèle créé à un temps ou un autre lieu. Enfin, il est aussi possible de restreindre spatialement le modèle afin d'obtenir une distribution d'espèce plus fidèle à la réalité (M-SDM).

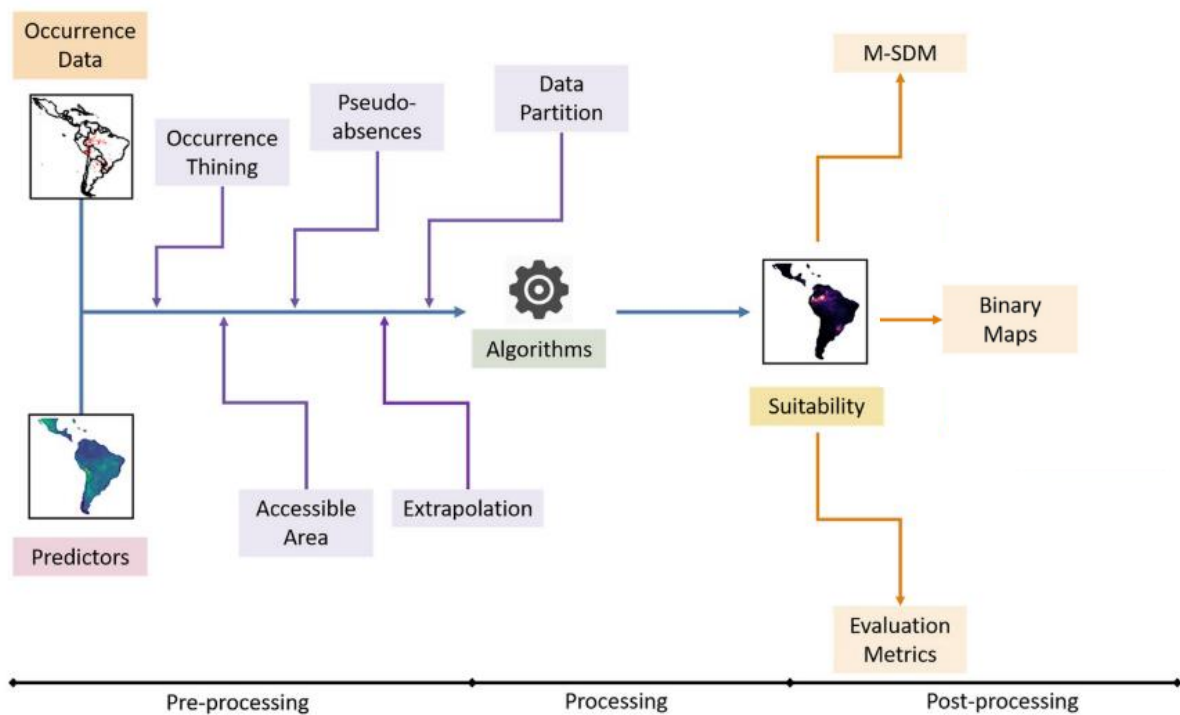


Figure 3 - Fonctionnement général de la construction d'un modèle de distribution d'espèce, modifié du processus du package ENMML de Andrade et al. (2020) à des fins d'illustration.

Choix d'un algorithme

Si les données d'entrée sont très importantes, c'est l'algorithme qui va générer le modèle. En fonction de l'objectif de l'étude et des données disponibles, il est donc essentiel de choisir le plus adapté.

Il existe plusieurs algorithmes en modélisation spatiale (Annexe 1), regroupés en différents types selon leur fonctionnement, et nombreuses sont les études qui les comparent (Elith and Leathwick, 2009; Franklin, 2010; Früh et al., 2018; Peterson, 2006b; Shabani et al., 2016).

Une première méthode est d'utiliser uniquement les données de présence, sans utiliser de jeu de données de comparaison. Ces méthodes sont dites '*true presence-only*', en opposition aux '*presence-only*' qui n'utilisent que les présences en entrée, mais génèrent des pseudo-absences ou un background. Cette méthode a été popularisée par l'algorithme BIOCLIM (Nix and Busby, 1986), qui utilisent les variables les plus contraignantes pour générer une prédiction. Cet algorithme peut aussi être utilisé avec des points d'absence pour générer une comparaison, ce qui le rend plus efficace. Une autre méthode de Machine Learning, la méthode '*One-Class Support Vector Machine*' est aussi utilisée pour les EEE (Drake and Bossenbroek, 2009; Guo et al., 2005). Cet algorithme de classification se focalise essentiellement sur la séparation entre la zone occupée et les autres, en utilisant uniquement les données de présence et les variables environnementales. Ces méthodes sont très dépendantes des données de présence, qui doivent impérativement être représentatives (Aikio et al., 2010). De plus, elles sont aussi très sensibles aux biais spatiaux (Aikio et al., 2010). Elles sont néanmoins plus simples à créer et à interpréter, mais souffrent d'un manque de précision et de réalisme (Elith, 2017).

Les autres algorithmes nécessitent un jeu de données de comparaison, soit en données d'entrée, soit par génération. On distingue les modèles de régression et les modèles de Machine Learning. Les premiers (GLM, GAM) fonctionnent essentiellement avec des données de vraie absence, utilisant des régressions pour distinguer les deux jeux de données. Les modèles de Machine Learning tels Boosted Regression Trees (BRT), Random Forest (RF), Maxent ou GARP sont des modèles de régression flexibles non paramétriques, qui vont améliorer le modèle par eux-mêmes, sans commande externe.

Pour une population à l'équilibre, les résultats des différentes méthodes varient fortement en fonction des données d'entraînement (Elith et al., 2006). D'une manière générale, les méthodes de régression, de classification et Maxent semblent très performantes, en comparant les résultats à l'écologie des espèces (Elith, 2017; Shabani et al., 2016). Concernant les modèles d'EEE, la comparaison est plus difficile car la distribution réelle de l'espèce est inconnue. Maxent, l'algorithme le plus répandu, semble néanmoins présenter des avantages. En effet, il est assez robuste quant aux problèmes de faible échantillonnage (Hernandez et al., 2006) et biais spatiaux (Graham et al., 2008; Loiselle et al., 2007). D'une manière générale, Maxent est considéré comme l'algorithme le plus efficace pour les EEE (Gormley et al., 2011), lorsque les données d'absence sont indisponibles.

Enfin, plusieurs études (Früh et al., 2018; Shabani et al., 2016) conseillent de tester plusieurs modèles équivalents (Maxent et GARP par exemple), et de les fusionner afin d'avoir une moyenne des résultats.

Réduction des biais d'échantillonnage

Dans le cadre de la modélisation des EEE, il existe plusieurs obstacles majeurs. Lorsque le modèle est dérivé d'une zone où l'espèce est présente depuis longtemps et en grande densité, il est commun de considérer que les données d'occurrence représentent suffisamment la distribution de l'espèce. Cependant, il existe plusieurs facteurs qui tendent à contredire cette hypothèse.

Premièrement, la densité des occurrences doit être représentative, ce qui n'est pas toujours le cas (Jarnevich et al., 2015). En effet, certaines régions sont plus souvent échantillonnées, ou profitent d'une détection plus avancée grâce notamment à l'essor de la télédétection. De cette manière, la différence de méthode d'échantillonnage d'une région à l'autre d'un même pays, ou d'un pays à un autre est parfois trop grande pour être négligée. Ainsi, il apparaît des biais d'échantillonnage, qui induisent en erreur l'algorithme. Deuxièmement, ce problème d'échantillonnage apparaît aussi localement. En effet, il est aisé de conclure que l'espèce aura été plus échantillonnée dans les zones dites accessibles (Barve et al., 2011; Kadmon et al., 2004). Les données d'occurrence favoriseront donc les bords de route, les orées de bois ou rives de cours d'eau, négligeant les nombreuses occurrences en zones moins fréquentées, ainsi que les propriétés privées, inaccessibles à l'échantillonnage.

Ces déclarations sont vraies pour les zones où l'espèce est présente depuis très longtemps, voire indigène. Elles le sont d'autant plus pour les régions récemment envahies. En effet, le taux d'échantillonnage est plus faible et certaines régions sont sur-échantillonnées tandis que d'autres ne le sont pas du tout. Les zones peu accessibles sont bien souvent négligées, ce qui implique un biais d'échantillonnage conséquent. A cela, il faut ajouter l'imprécision géographique et taxonomique des données ainsi que le faible nombre d'occurrence. Bien que n'impactant en rien le caractère biaisé de l'échantillon, ceux-ci contribuent à fragiliser le jeu de données. Celles-ci doivent donc être l'objet d'une attention particulière dans le pré-processing et l'interprétation des résultats.

Il existe plusieurs manières de réduire ces biais d'échantillonnage:

- Le filtrage spatial des données d'occurrence (Boria et al., 2014; Warren et al., 2014)
- La manipulation de la génération des pseudo-absences (Kramer-Schadt et al., 2013)
- L'intégration d'une variable supplémentaire dans l'algorithme (Václavík et al., 2012)

D'une manière générale, un premier filtrage spatial est toujours réalisé, afin de réduire la différence de densité spatiale des données. Mais afin de réduire au maximum les biais d'échantillonnage, il peut être nécessaire de filtrer davantage, afin de minimiser les faux positifs (erreurs de commission ; Rondinini et al. (2006)) et les faux négatifs (erreurs d'omission ; Rondinini et al. (2006)). Si l'échantillon est trop faible, ou que le biais est trop grand, la manipulation des pseudo-absences peut aussi réduire le biais. Dans ce cas, il est donc impératif de choisir un algorithme nécessitant des pseudo-absences.

Avant toute modélisation, étant donné que les biais d'échantillonnage sont des problèmes couramment rencontrés dans la modélisation d'aires de répartition d'EEE, il peut être intéressant de constater les différentes méthodes utilisées dans la littérature.

Le filtrage spatial

Pour réduire les biais d'échantillonnage, la première chose à faire est un filtrage spatial (Anderson and Raza, 2010; Boria et al., 2014). Dans les cas de biais relativement faibles, le filtrage suffit pour générer des modèles satisfaisants. Il existe plusieurs techniques de filtrage, toutes basées sur la réduction de la densité des points d'occurrence.

Premièrement, il est préférable de ne garder qu'une donnée d'occurrence par cellule. Dans les cas où les données d'occurrence sont rares, la manipulation de background peut aider à réduire les biais. Si

le nombre de points de présence est suffisant, il est possible de réduire la densité en supprimant aléatoirement les points trop proches. La distance est à fixer en fonction de l'espèce et de l'étendue/la géographie de la zone d'étude. Cette distance doit être choisie pour refléter le biais géographique lié à l'échantillonnage (et donc à l'hétérogénéité du paysage), plus que la dispersion de l'espèce.

Cette technique appauvrit le jeu de données, mais permet de réduire considérablement les biais d'échantillonnage. Il est important de noter que la réduction du jeu de données elle-même peut améliorer la qualité du modèle, limitant l'impact des biais indirectement (Boria et al., 2014).

La manipulation de la génération des pseudo-absences

Lorsque le filtrage spatial ne suffit pas, ou que le nombre d'occurrence n'est pas assez grand pour le permettre, il est préférable de manipuler la génération des pseudo-absences durant le pré-processing. En effet, les pseudo-absences permettant une comparaison entre deux jeux de données, une correction peut être apportée si ceux-ci sont biaisés de la même manière. Ces méthodes diminuent le risque d'erreur d'omission pour les espèces à l'amplitude écologique étroite, mais est moins adaptée aux espèces généralistes, pour qui elle risque de l'augmenter (Kramer-Schadt et al., 2013). Ces méthodes seront énumérées dans la section suivante consacrée à la génération des points de pseudo-absences.

Autres points d'attention

Lorsque l'on aborde la distribution spatiale des EEE, il est important de prendre en considération la 'Spatial Auto-Correlation' (SAC). Celle-ci représente la structure particulière d'une espèce dans l'espace. En effet, il existe une autocorrélation spatiale très forte dans les populations lors de la colonisation (Václavík et al., 2012). Une espèce est introduite à un endroit, puis s'y développe, ce qui résulte en une plus forte densité de cette espèce dans cette région. De cette manière, certaines données d'occurrence ne sont pas dues à des biais d'échantillonnage, mais plutôt à cette structure spatiale propre à l'espèce lors de sa colonisation. Intégrer cette structure dans le modèle peut par exemple être un moyen d'intégrer la dynamique de l'espèce, sans recourir à un modèle dynamique, rarement réalisable. Il est par exemple possible d'ajouter une variable de structure spatiale lors d'une modélisation de distribution de l'espèce, afin de mimer la dispersion spatiale (Václavík et al., 2012).

Distinguer les SAC et les biais d'échantillonnage est donc important lorsque l'on travaille avec une EEE, spécialement lorsqu'il s'agit d'une espèce nouvellement introduite ou à un stade précoce d'invasion. Parmi les algorithmes fréquemment utilisés, Maxent semble être le moins sensible aux SAC. (Václavík et al., 2012).

Génération des pseudo-absences

Lors de la phase de pré-processing, il n'y a pas beaucoup de facteurs à modifier pour optimiser le modèle. La génération des pseudo-absences, dans le cas des modèles '*presence-only*', est une étape cruciale. En effet, le modèle est entièrement tributaire des données en entrée, et d'autant plus de celles créées de toute pièce. Ces points générés serviront de comparaison pour identifier les zones favorables ou non à une espèce. Mais cette génération, aléatoire ou non, présente plusieurs failles. D'une part, lorsqu'elle est aléatoire, il est très probable d'inclure des fausses absences dans le modèle (Gu and Swihart, 2004). Ce genre d'erreurs, si elles sont fréquentes, peuvent grandement affecter le modèle et provoquer de mauvaises estimations (Chefaoui and Lobo, 2008). Ainsi, il est préférable de travailler avec des vrais points d'absence lorsqu'ils sont disponibles.

En effet, elles réduisent la part d'aléatoire dans le modèle, ce qui est toujours une bonne chose. Il existe cependant beaucoup de critiques sur l'utilisation des absences, notamment sur plusieurs facteurs tels que les interactions biotiques, les contraintes de dispersion ou les perturbations, qui

affecteraient les absences, les rendant peu fiables (Jiménez-Valverde et al., 2008). Cependant, il est à noter que ces facteurs influent aussi sur les présences, rendant ces arguments peu valides (Elith, 2017; Elith et al., 2011). Le problème majeur en travaillant avec des vraies données d'absence sont les fausses absences (Jiménez-Valverde et al., 2008), mais ce problème reste présent, voire plus conséquent avec un background aléatoire, tant est que les données d'absence soient suffisamment nombreuses. Enfin, travailler avec des données d'absence reste problématique si l'échantillonnage est effectué différemment pour les présences/absences, par exemple en extrayant les données Gbif pour ajouter des présences supplémentaires dans le jeu de données. Il faut alors recourir à différentes méthodes pour réduire les biais d'échantillonnage, d'autant qu'il existe un biais associé aux présences, et l'autre aux absences.

Dans la majorité des cas cependant, les données d'absence ne sont pas accessibles et un background est créé pour opérer la comparaison. Il est fortement recommandé d'utiliser un algorithme adapté aux données disponibles. Un algorithme qui utilise des données d'absence et de présence en entrée sera en effet moins performant si les données d'absence sont créées artificiellement. De la même manière, utiliser des vraies données d'absence comme données de background pour un algorithme de type 'presence-only' ne donnera pas de meilleurs résultats.

Pour une grande partie des modèles, les pseudo-absences sont générées aléatoirement sur l'étendue de l'étude, à hauteur de 10 000 points. Ces paramètres peuvent toutefois être modifiés, afin de mieux correspondre à l'objectif ou pour réduire un problème rencontré, par exemple les biais d'échantillonnage.

Enfin, dans le cadre des modèles traitant des EEE, des extrapolations de modèles (dans l'espace ou dans le temps) sont souvent effectuées. Dans ces conditions, la génération de pseudo-absences doit être l'objet d'une attention particulière. En effet, le background peut être différent d'un endroit à l'autre, faussant le modèle extrapolé. Un background inapproprié dans le modèle de départ peut grandement affecter les extrapolations spatiales (Mau-Crimmins et al., 2006; Steiner et al., 2008) et temporelles (Hilbert, 2001).

Généralement, il existe trois variables dans la génération du background :

- Le nombre de points générés
- La position et la manière de les générer
- L'étendue sur laquelle les générer

Nombre de points

Le nombre de points constituant le background du modèle est un sujet qui divise. Il a été prouvé que le ratio présences/absences influait sur la précision du modèle (McPHERSON et al., 2004), et il est donc essentiel de s'interroger sur celui-ci. Certains algorithmes, dont Maxent, créent automatiquement 10 000 points de pseudo-absence, nombre généralement retenus pour avoir la meilleure précision (Phillips and Dudík, 2008).

Cependant, certaines études ont montré qu'un même nombre d'absences que de présences était favorable à une meilleure précision dans certains cas. De cette manière, les méthodes de régressions (GLM, GAM) se font généralement avec un ratio présences/pseudo-absences de 1 :1, tandis que les autres méthodes utiliseront 10 000 pseudo-absences (Barbet-Massin et al., 2012; Phillips and Dudík, 2008) afin de générer un background. Enfin, le ratio 1 :10 est parfois utilisé lorsque l'étendue de la zone d'étude est très restreinte ou au contraire très large. Ce ratio correspond à la limite au-delà de laquelle la précision du modèle tend à stagner pour plusieurs algorithmes (BRT, RF) (Barbet-Massin et al., 2012).

Encore maintenant, un consensus ne semble pas avoir été trouvé, même si la majorité des études se contente de prendre la valeur par défaut des algorithmes de Machine Learning, à savoir 10 000 points de pseudo-absence.

Etendue utilisée

La grande majorité des modèles utilise la même étendue pour le background que pour la zone d'étude. Cependant, certaines études se sont intéressées à la question (Lobo et al., 2010; Phillips and Dudík, 2008; Zaniwski et al., 2002). Si les données environnementales le permettent, il peut être favorable d'étendre le background au-delà de la zone d'étude (VanDerWal et al., 2009). Cette affirmation n'est vraie que lorsqu'il n'existe pas (ou très peu) de données de présence à l'extérieur de la zone d'étude. Cela peut être très bénéfique au modèle, particulièrement si les environs de la zone d'étude comportent de vastes étendues non favorables à l'espèce.

Dans ce cas, il est toutefois important de se rappeler que les données d'absence ne doivent pas être trop éloignées des données de présence, au risque de donner des surestimations dans le modèle (Lobo et al., 2010). Dans le cadre des EEE, ce type de facteur est rarement examiné, étant donné que les EEE sont essentiellement généralistes, et colonisent donc souvent plusieurs types de milieux.

Manière de générer les points

La méthode de génération des points de pseudo-absence peut répondre à deux problématiques. D'une part, la réduction des biais d'échantillonnage, abordée à la section précédente, et d'autre part la réduction des fausses absences.

Pour réduire les biais d'échantillonnage

Il existe beaucoup de méthodes de manipulation de background ou des pseudo-absences, la plus connue étant la '*Target Group Selection*' ou '*Target Group Background*', consistant à prendre comme donnée de background les données d'occurrence d'une espèce similaire (Phillips et al., 2009). Il est dans ce cas nécessaire de choisir une espèce taxonomiquement proche échantillonnée de la même manière, afin de répliquer le biais dans le background. Bien que cette méthode soit largement répandue dans la réduction de biais, elle est aussi beaucoup critiquée. Elle présente de meilleurs AUC (indice de validation largement répandu en modélisation, mais aussi très critiqué ; voir paragraphe sur les indices de validation), mais le modèle est moins bien calibré et il sous-estime grandement la distribution. En effet, certaines variables peuvent être écartées parce qu'elles sont essentielles aux deux espèces et certaines régions géographiques peuvent être très biaisées par ce background. L'hypothèse la plus souvent retenue est que l'autocorrélation spatiale est plus grande avec cette méthode, ce qui renforce l'AUC. De plus, dans le cas où le biais serait faible, le modèle sera davantage une distinction entre les deux espèces plutôt qu'un modèle de distribution potentielle (Warren et al., 2014). Enfin, il peut y avoir de nombreuses erreurs d'omission dans le cas d'espèces à l'amplitude écologique large. Les EEE colonisant souvent plusieurs types de milieux, cette méthode n'est pas recommandée.

Une deuxième technique utilisée est la stratification du background (Barbet-Massin et al., 2012). En effet, si le biais est lié à un sur-échantillonnage d'un habitat particulier, il est possible de fractionner le background pour avoir un même nombre de points dans chaque type d'habitat défini (Albert et al., 2010; Kadmon et al., 2004). Il est possible de faire une division équivalente pour les données climatiques (Barbet-Massin et al., 2010). Ces techniques sont toutefois à éviter, étant fortement tributaire du type de biais rencontré et peu efficaces (Barbet-Massin et al., 2012).

D'autres approches plus mécanistes existent. Ainsi, les points de pseudo-absence peuvent être sélectionnés uniquement dans les endroits accessibles, en région non favorable à l'espèce (Chapman

et al., 2019). Ceci demande des données supplémentaires et une connaissance élevée de l'espèce. De surcroît, ce type de modèle contraint souvent la sélection de variables, étant donné que les zones non favorables sont sélectionnées manuellement. Cependant, ces modèles répondent aux critiques reprochant aux modèles corrélatifs de ne pas intégrer la biologie de l'espèce.

Il est difficile de conclure quelle méthode est la meilleure, le résultat dépendant largement de l'espèce, de la région et du type de biais rencontré.

Pour réduire les fausses absences

Il existe plusieurs manières de générer les pseudo-absences, afin de tenir compte du chevauchement des points, de l'environnement ou encore du climat. Plusieurs études ont comparés différentes techniques (Chefaoui and Lobo, 2008; Lütolf et al., 2006; Phillips and Dudík, 2008).

- Background aléatoire total, largement utilisé en littérature car généré automatiquement par la plupart des algorithmes.
- Background aléatoire avec des zones d'exclusion autour des points d'occurrence, afin d'éviter le chevauchement des différents jeux de données. Le rayon est habituellement déterminé en fonction de l'espèce étudiée (Chefaoui and Lobo, 2008).
- Background aléatoire avec zones d'exclusion autour des points d'occurrence, et avec des zones d'inclusion autour de ces mêmes points, pour un rayon plus grand. Le but de ce background est de limiter les fausses absences fréquentes lors d'un échantillonnage incomplet, en supposant la zone d'inclusion correctement échantillonnée.
- Background situé sur les sites historiquement connus pour ne pas avoir hébergé l'espèce. Ces sites sont considérés comme non favorables. Ce type de background n'est pas applicable aux espèces en début de colonisation.
- Background situé dans les sites où l'espèce n'est pas présente, mais d'un habitat similaire. Ce type de background donne des résultats concluants, mais demande une grande ressource de données (Lütolf et al., 2006).
- Enfin, d'autres techniques consistent à utiliser des algorithmes spécifiques pour générer un background. La première étape est d'utiliser un algorithme 'true presence only' tels BIOCLIM ou One-Class Support Vector Machines (OCSVM) afin d'identifier sommairement les régions les moins favorables à l'espèce. Ensuite, une génération aléatoire de points de pseudo-absence sur cette zone est effectuée pour la modélisation finale. Cette méthode est appelée 'Random Selection with environmental profiling' (RSEP) par (Iturbide et al., 2015). Celle-ci permet de réduire les fausses absences, et peut être améliorée en y ajoutant deux étapes, via la 'Two-Step Method' (TS). La deuxième étape est de générer plusieurs modèles avec la distance des points de pseudo-absence aux données de présence comme variable (Figure 4). Ces modèles sont donc construits avec des étendues de background différentes, lui-même

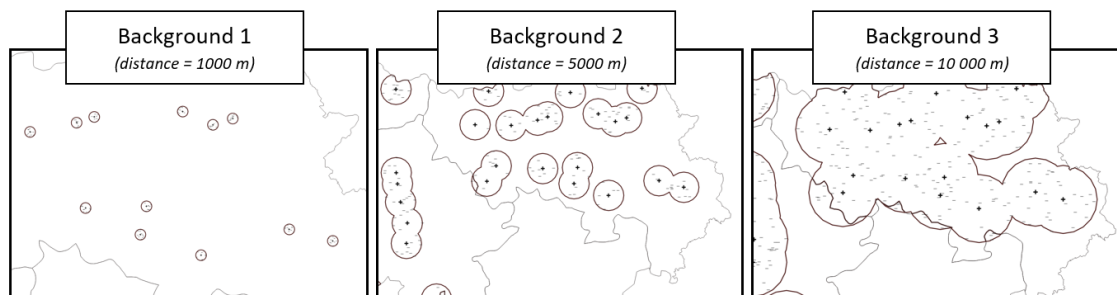


Figure 4 – Exemples de différentes distances aux présences pour la génération du background. Les '+' représentent les points de présence tandis que les '-' représentent les pseudo-absences générées. Lors de la 'Two-Step Method' (TS), plusieurs distances sont testées et comparées via des indices métriques afin de choisir la meilleure distance.

généralisé sur les zones identifiées par la première étape. Enfin, la troisième étape consiste dans le calcul de différents indices métriques afin de choisir le meilleur modèle. Ces deux méthodes produisent des modèles de bonne qualité, (Iturbide et al., 2015) recommandant la TS comme méthode la plus efficace bien que la différence de qualité soit faible (Iturbide et al., 2015).

La meilleure méthode dépend de plusieurs facteurs. D'une part, l'espèce étudiée et sa dispersion, mais aussi le type d'échantillonnage qui a été effectué (Fourcade et al., 2014). D'autre part, si l'espèce n'est pas en équilibre dans son environnement, il est préférable de sélectionner aléatoirement (Chefaoui and Lobo, 2008).

Validation du modèle

L'élaboration du modèle en elle-même a pour but d'être la plus précise possible. Mais étant donné que les données d'occurrence sont incomplètes, il est nécessaire de calculer et d'interpréter un ou des indices métriques d'évaluation. La méthode utilisée pour évaluer un modèle peut varier fortement d'une étude à l'autre, et doit être l'objet d'une attention toute particulière. En effet, les résultats de cette étape sont les arguments qui vont donner du poids au modèle.

La validation par calibration-évaluation

La méthode la plus répandue est de sélectionner une fraction du jeu de données pour évaluer la qualité du modèle. De cette manière, le jeu de données de départ est divisé en jeu de calibration et jeu d'évaluation, dans une proportion qui varie de 3 :1 à 5 :1. Il est important que le modèle soit calibré uniquement avec le jeu de calibration à défaut de surestimer la qualité du modèle. De plus, ce jeu de données d'évaluation doit, dans la mesure du possible, être indépendant du jeu de données de calibration (Radosavljevic and Anderson, 2014). En effet, les deux jeux de données, si sélectionnés aléatoirement, provoquent quelques problèmes (Araujo et al., 2005). Premièrement, en sélectionnant aléatoirement, les points de calibration et d'évaluation sont souvent proches les uns des autres. Ceci résulte en une autocorrélation spatiale qui surestime l'évaluation (Hijmans, 2012). De plus, les biais géographiques présents dans le jeu de calibration sont aussi présents dans le jeu d'évaluation, ce qui empêche toute estimation du biais et toute évaluation des performances réelles du modèle. Ainsi idéalement, le jeu d'évaluation devrait totalement être indépendant, en venant par exemple d'une autre zone géographique/temporelle (Radosavljevic and Anderson, 2014). Les données temporelles étant souvent plus rares, il est donc recommandé de prendre un jeu d'évaluation d'un espace donné de la zone d'étude. Cependant, il est à noter que les données de background utilisées ne doivent donc être générées que sur la zone de calibration (Anderson and Raza, 2010).

Bien que cette méthode soit plus efficace, elle est peu répandue car plus longue à mettre en œuvre.

Lorsque le nombre d'occurrence permet difficilement de laisser de côté une partie du jeu de données initial pour calibrer le modèle, il est possible (et fréquent) de faire une validation croisée (Hijmans, 2012). Le jeu de données est alors fractionné en k jeux. k modèles sont créés, avec à chaque fois un des sous-jeux différent servant de jeu d'évaluation. $k-1$ jeux de données sont alors utilisés comme jeu de calibration. De cette manière, chaque modèle produit une évaluation différente. Une analyse de la moyenne et de l'écart-type des indices métriques d'évaluation de ces modèles peut alors être réalisée. Il est possible de coupler la validation croisée et le fractionnement spatial en divisant spatialement la zone d'étude en k zones de même nombre d'échantillon (Radosavljevic and Anderson, 2014).

L'aire sous la courbe ROC

L'indice le plus utilisé et sûrement le plus controversé (Lobo et al., 2008) est l'aire sous la courbe (AUC) de la fonction d'efficacité du récepteur ou courbe 'receiver operating characteristic' (ROC) (figure 5). La valeur de l'AUC correspond à la probabilité que le modèle classe correctement un point, et varie de 0 à 1, la valeur la plus haute étant la plus performante pour classer les données.

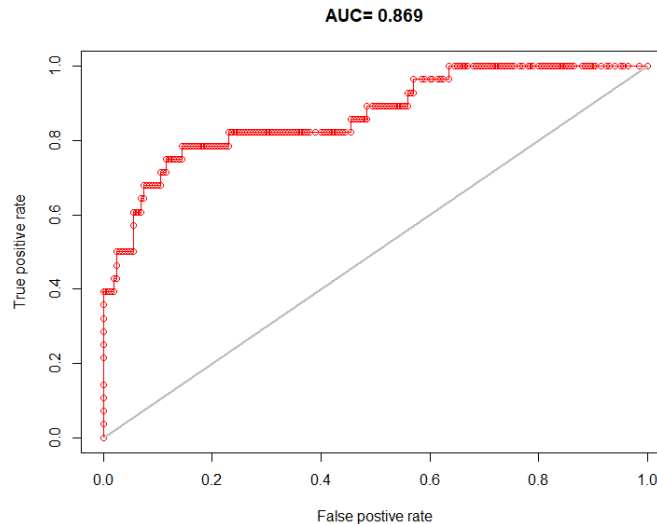


Figure 5 – Exemple de courbe ROC et de valeur AUC correspondante. Le taux de faux positifs est en abscisse et le taux de vrais positifs est en ordonnée. La courbe ROC est en rouge et constituée des différentes valeurs obtenues pour différents seuils de classification, chacune représentée par un rond rouge. La ligne diagonale représente la limite sous laquelle la courbe traduit une classification hasardeuse. Enfin, l'AUC correspond à l'aire sous la courbe ROC.

L'abscisse de la courbe ROC est le taux de faux positifs (Equation 3) et l'ordonnée est le taux de vrais positifs (Equation 1). Ces taux sont calculés pour différents seuils de classification. Ainsi, pour un modèle, la ROC va tester des dizaines de seuils différents afin de classer correctement les points en zone favorable/défavorable. Ces résultats sont reliés afin de créer la courbe ROC. Plus le modèle est bon, plus la courbe maximisera l'aire sous la courbe, car la valeur du seuil *réel* sera plus précise.

Le seuil de présence/absence ou défavorable/favorable selon l'objectif de l'étude, est calculé comme la valeur du modèle qui maximise l'abscisse et l'ordonnée de la ROC. Le taux de vrais positifs est alors appelé sensibilité (Equation 1), et le taux de vrais négatifs est appelé spécificité (Equation 2).

Équation 1 - taux de vrais positifs ou sensibilité

$$\text{True Positive Rate/Sensitivity} = \frac{\text{True positive}}{\text{True positive} + \text{False negative}}$$

Équation 2 - Spécificité

$$\text{Specificity} = \frac{\text{True negative}}{\text{True negative} + \text{False positive}}$$

Équation 3 - Taux de faux positifs

$$\text{False Positive Rate} = \frac{\text{False positive}}{\text{True negative} + \text{False positive}} = 1 - \text{Specificity}$$

Cet indice est très controversé, pour plusieurs raisons (Lobo et al., 2008). L'AUC donne la même importance aux erreurs d'omission et de commission, or les données d'absence sont générées, ce qui augmente drastiquement les erreurs d'omission. De plus, aucune composante spatiale n'est prise en compte dans l'indice, ce qui fait de l'indice AUC un indice peu adapté aux modélisations spatiales. Les valeurs pour la spécificité étant très variables d'un modèle à l'autre, l'AUC est dès lors très déconseillé pour comparer des modèles de différentes zones/espèces entre eux. De plus, l'AUC est connu comme étant très sensible aux biais d'échantillonnage, défavorisant les modèles aux biais moindres (Hijmans, 2012; Phillips et al., 2009).

De plus, il donne des valeurs très faibles aux espèces généralistes, bien que les modèles soient précis. L'AUC ne prodigue donc pas une bonne information sur la performance du modèle, en particulier pour les modèles de EEE, souvent généralistes, aux données incomplètes et dont l'échantillonnage est très sensible aux biais.

Néanmoins, l'indice est encore largement utilisé dans la littérature, par mesure d'habitude et de facilité. De fait, il peut donner une bonne estimation et produire un seuil qui est objectif (Hijmans, 2012), mais ne doit pas être la seule méthode d'évaluation, au risque d'être biaisé.

L'AICc

Un autre indice couramment utilisé est le critère d'information d'Akaike (AIC ; Equation 1), régulièrement employé dans sa forme corrigée pour petits échantillons (AICc ; Equation 5).

Equation 4 - Critère d'information d'Akaike, où k est le nombre de paramètres à estimer du modèle et L est le maximum de la fonction de vraisemblance du modèle.

$$AIC = 2k - 2\ln(L)$$

Equation 5 - Critère d'information d'Akaike pour les petits échantillons, où n est la taille de l'échantillon

$$AICc = AIC + \frac{2k(k + 1)}{n - k - 1}$$

Le modèle choisi sera alors celui qui minimise l'AICc. L'indice est un compromis entre la qualité de l'ajustement et la complexité du modèle, favorisant les modèles avec moins de variables, ce qui réduit le surajustement. Les résultats de Maxent dépendent fortement de la complexité du modèle (Anderson and Gonzalez, 2011; Elith et al., 2010; Phillips and Dudík, 2008; Warren and Seifert, 2011), c'est donc un facteur important à prendre en compte (Radosavljevic and Anderson, 2014).

Les modèles moins complexes étant reconnus comme favorables à l'extrapolation (Warren et al., 2014; Warren and Seifert, 2011), cet indice est largement utilisé en étude des EEE. Cependant, un trop faible nombre de variable est aussi préjudiciable à la qualité du modèle, il est donc important de ne pas systématiquement prendre le modèle à l'AICc le plus bas, souvent pauvre en variables, ce qui surestimera la distribution.

Autres indices

Outre l'AUC et l'AICc, il existe plusieurs autres indices, bien que moins répandus. Le TSS par exemple (sensibilité + spécificité -1), très fortement corrélé à l'AUC (Allouche et al., 2006), mais plus aisé à calculer de par sa simplicité. Facilement interprétable, il souffre néanmoins des mêmes défauts que l'AUC.

Un autre indice fortement utilisé est l'indice Kappa de Cohen qui compare la prédiction du modèle avec une classification aléatoire. Il varie de -1 à 1, une valeur de 0 ou négative montrant un résultat comparable à l'aléatoire. L'indice est prisé pour sa simplicité son apport d'information différente, mais

a été reconnu comme peu adapté à la modélisation spatiale (Allouche et al., 2006). Tout comme l'AUC, l'indice semble inadéquat pour comparer différents modèles dans le temps/espace/différentes espèces (McPHERSON et al., 2004; Warren et al., 2014), ce qui en fait un indice peu utile en science des EEE.

La validation sur le terrain

Enfin, même si ces indices statistiques peuvent parfois donner une bonne estimation de la qualité d'une carte de prédiction, ils ne remplacent pas une validation sur le terrain, seule véritable évaluation qui confirmera/infirmera le modèle (Soberón, 2010; West et al., 2016). Une telle opération est bien souvent impossible, mais il peut être utile de réaliser un échantillonnage en zone favorable et en zone défavorable pour constater sur le terrain des résultats du modèle.

Dans le même état d'esprit, plusieurs modèles servent ainsi à guider un échantillonnage plus rigoureux, permettant ainsi des données plus complètes. De cette manière, plusieurs modèles sont réalisés après chaque échantillonnage, guidés par les zones de présence identifiées par les modèles précédents (Jarnevich et al., 2015).

Les indices de validation pour les EEE

Il est donc très compliqué d'évaluer un modèle, en particulier pour les modèles traitant d'EEE. Dans la majorité des cas, une extrapolation d'un modèle est faite pour déduire la distribution potentielle d'une EEE dans une nouvelle région. Comme expliqué précédemment, certains indices sont peu adaptés à cette contrainte. Il est donc courant dans la littérature d'évaluer la qualité d'un modèle par les variables qui ont été retenues, par comparaison avec un modèle indépendant ou simplement par expertise d'un écologue. Dans le cas d'une extrapolation, l'évaluation ne se fait pas sur la zone d'étude, et concerne donc le modèle dans une autre zone géographique, ce qui peut poser parfois problème.

Les données des zones envahies, rarement à l'équilibre, sont problématiques pour évaluer le modèle. En effet, ces données comprennent des présences en zone défavorable de par la dynamique de dispersion des EEE, ce qui peut nuire à l'évaluation. De cette manière, malgré une bonne performance du modèle, les indices métriques seront faibles si le jeu d'évaluation comprend trop de points d'occurrence présents en zone défavorable (Elith, 2017).

Il faut donc manier avec prudence ces indices d'évaluation, particulièrement lors d'une étude sur les espèces exotiques envahissantes.

Matériel et méthode

L'espèce étudiée : *Ambrosia artemisiifolia* L.

Biologie de l'espèce

L'ambrosie à feuilles d'armoise, ambrosie annuelle ou *Ambrosia artemisiifolia* L. est une astéracée annuelle (Figure 6), originaire d'Amérique du Nord. Plus précisément, la zone d'indigénat de l'espèce se situe aux Etats-Unis et au sud du Canada (Bassett and Crompton, 1975), avec possiblement un élargissement au nord du Mexique. Cependant, son caractère invasif lui a permis de coloniser tous les continents.

Elle peut montrer une variation phénotypique très grande, tant du point de vue de la taille (Ortmans et al., 2016a) que des graines (Fumanal et al., 2007; Ortmans et al., 2016b). Elle peut mesurer de 10 cm à 2,50 m en fonction des conditions environnementales (Essl et al., 2015) et a une racine pivotante qui peut aller jusque quatre mètres de long.



Figure 6 - Illustration d' *Ambrosia artemisiifolia* à son stade végétatif (A) et en fleur (B)

En Europe, elle atteint son optimum dans les climats tempérés, caractérisés par des hautes températures estivales durant sa période de croissance (Essl et al., 2009). De fait, le facteur limitant le plus sa colonisation serait la température moyenne des mois chauds. Cela expliquerait l'absence de l'espèce dans les endroits à l'altitude élevée, caractérisé par des températures estivales plus faibles. Au sud, l'espèce serait contrainte par les précipitations estivales trop faibles (Chapman et al., 2014), et les températures hivernales trop hautes (Shrestha et al., 1999).

L'espèce a une niche écologique très large. Elle s'installe sur des sols riches ou rudéraux principalement. Elle ne semble pas avoir de préférence bien marquée au niveau du sol (B Fumanal et al., 2008) bien qu'elle semble préférer les sols argileux et sableux de l'Amérique du Nord (Bassett and Crompton, 1975). Différents auteurs se contredisent concernant l'acidité idéale, allant d'un pH acide (Gusev, 2019) à basique (Landolt et al., 2010). Elle n'a pas de préférence marquée pour le taux d'azote, qui selon les auteurs peut aller de *faible* à *élevé* (Gusev, 2019). Malgré ces divergences, tous s'accordent à dire que ces deux facteurs ne limitent pas la dispersion de l'espèce. L'espèce a par contre une nette préférence pour les sols secs, et est relativement exigeante quant à la lumière, ne poussant qu'à une certaine intensité de lumière (Gusev, 2019).

L'espèce colonise les champs arables et les bords de route dans les milieux fortement envahis (Essl et al., 2009; Milakovic et al., 2014; Pinke et al., 2013). En début de colonisation, elle touche essentiellement les milieux rudéraux (Bullock et al., 2012). Enfin, elle est peu compétitrice, et ne colonise les milieux fermés qu'après une perturbation (Bazzaz, 1974; Brandes and Nitzsche, 2006). La taille des populations peut fortement varier en étendue et en densité. Les régions peu envahies se caractérisent par des petites populations éphémères, mais les régions fortement envahies peuvent accueillir de grandes populations de plusieurs millions de plants, parfois très denses (jusqu'à 400 plants/m²) (Bullock et al., 2012).

La production de graine serait liée à la biomasse du plant (Dickerson and Tolbert, 1968; Fumanal et al., 2007), et donc à la compétition et à l'habitat. Le nombre de graines peut aller de 300 à 6000 par plante, voire jusque 14 000 quand les conditions sont optimales (Bassett and Crompton, 1975). La dispersion naturelle des graines se fait essentiellement par anémochorie et barochorie, à une distance maximale de deux mètres de la plante mère (Dickerson and Tolbert, 1968), les akènes n'étant pas adaptés pour une dispersion sur des longues distances. Quelques cas de zoochorie par des oiseaux ou des rongeurs (Nitzsche, 2010) ont toutefois été relevés, mais sont d'une importance mineure. La dispersion à grande distance s'explique essentiellement par des activités humaines, tels que la contamination des semences et des graines à destination animale, les machines agricoles et le déplacement des sols contaminés (Bullock et al., 2012). Ainsi, les lieux d'introduction primaires sont essentiellement les champs de tournesols, de betteraves sucrières, de maïs ou de soja, ou encore les poulaillers. En effet, les graines à destination animale sont connues pour être grandement contaminées (Brandes and Nitzsche, 2006; Chapman et al., 2016).

La dormance de l'ambrosie à feuilles d'armoise est interrompue par une longue période froide (Essl et al., 2015; Willemsen, 1975), facilitée par peu de compétition ou l'arrivée d'une perturbation (Bazzaz, 1979; B Fumanal et al., 2008). Les graines qui ne germent pas entrent alors dans une dormance secondaire (Baskin and Baskin, 1980; Bazzaz, 1979; Willemsen, 1975), provoquée par une luminosité trop faible, une faible variation de température ou un été trop chaud ou trop sec (Bazzaz, 1979). Cette seconde dormance qui alimente la banque de graine est très problématique pour la lutte contre l'espèce (Baskin and Baskin, 1980; Guillemain and Chauvel, 2011). Très résistante, elle peut dépasser les 1000 graines au mètre cube. Ces graines peuvent rester viables jusque 40 ans dans les meilleures conditions (Bassett and Crompton, 1975; Fumanal et al., 2007), ce qui joue un rôle majeur dans son invasion. Dans les champs cultivés, cela pose évidemment problème à cause du labour qui enterre et recouvre les graines (Baskin and Baskin, 1980; Bassett and Crompton, 1975; B. Fumanal et al., 2008; Guillemain and Chauvel, 2011). De plus, cela peut induire en erreur les observations de l'espèce. En effet, il est aisé de conclure que l'espèce est naturalisée, alors qu'il peut s'agir d'une expression de la banque de graines sur plusieurs années, sans aucune production de graine récente.

Les graines émergent en été, avec un taux de croissance de 80% en moyenne (Milakovic and Karrer, 2016) mais qui dépend fortement de la température (Essl et al., 2015). Dans les climats tempérés, les fleurs mâles et femelles sont produites de l'été à fin automne (juillet à octobre) (Essl et al., 2015).

Le tournesol et l'ambrosie étant de la même famille botanique des astéracées, le contrôle chimique de l'espèce pose problème (Chollet et al., 1999). Malgré la haute résistance aux herbicides de certaines variétés de tournesol, une très grande densité d'ambrosie peut tout de même être observée dans les champs traités (B Fumanal et al., 2008; Genton et al., 2005; Ozaslan et al., 2016; Pinke et al., 2013). De plus, elle possède une très grande variabilité génétique dans les populations colonisatrices (B Fumanal et al., 2008; Milakovic and Karrer, 2016).

L'espèce a été introduite en Europe il y a plus d'un siècle ((Heckel, 1906)), et a envahi de nombreux pays européens (Chauvel et al., 2006; Kazinczi et al., 2008; Smith et al., 2013) où elle cause de grands soucis sanitaires à cause de son pollen très allergénique (Kazinczi et al., 2008; Smith et al., 2013), répandu d'août à septembre. Les symptômes sont divers, allant de la simple gêne respiratoire jusqu'à des fortes crises d'asthme, ainsi que des conjonctivites ou des réactions cutanées. La situation sanitaire en Europe est préoccupante, de par le nombre croissant de personnes sensibles au pollen d'ambroisie. En effet, les quantités de pollen émises dans l'atmosphères par l'ambroisie en Europe sont en constante augmentation (Fumanal, 2007) (Figure 7).

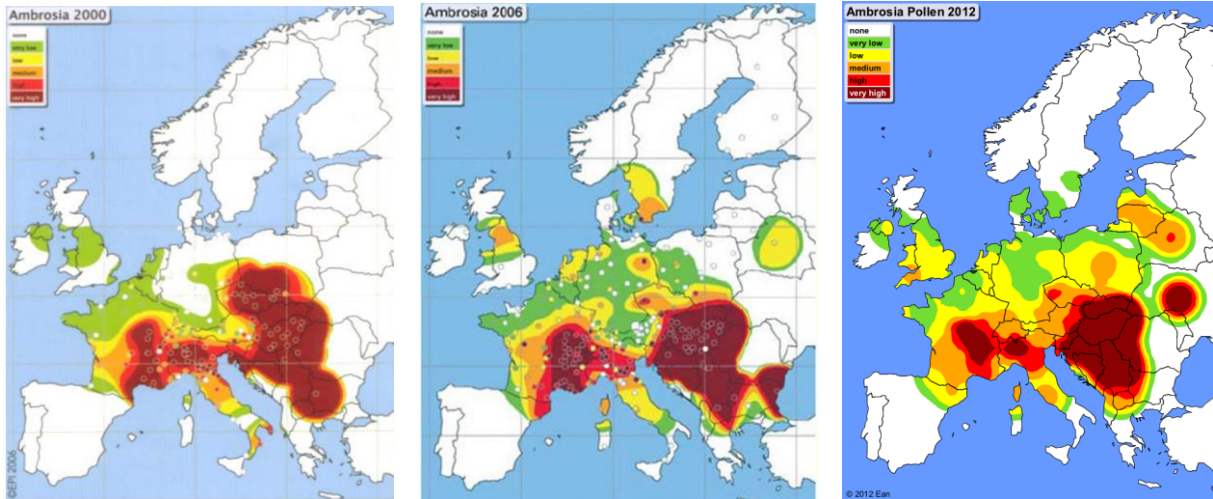


Figure 7 - Quantité de pollen d'*Ambrosia artemisiifolia* émise dans l'atmosphère entre 2000 et 2012 en Europe. Données issues de l'European Aeroallergen Network (EAN) et de l'European Pollen Information Ltd. (EPI).

L'espèce étant particulièrement nuisible dû à son pouvoir allergénique et sa capacité de ravageuse de culture, elle figure parmi les EEE les mieux étudiées.

Répartition de l'espèce dans le monde

Ambrosia artemisiifolia a envahi tous les continents, particulièrement l'Europe et l'Asie. A ce jour, les pays non-européens les plus impactés sont l'Australie, la Chine (Chen et al., 2007; Qin et al., 2014), l'Inde, le Japon, la péninsule de Corée, l'Azerbaïdjan, la Russie (Reznik, 2009), la Turquie, la Géorgie et l'île d'Hawaï, où l'espèce est considérée comme invasive.

Il est assez difficile de conclure du statut d'espèce exotique pour les pays suivants : l'Argentine, les Bahamas, le Brésil, la Colombie, Cuba, l'Equateur, l'île d'Hispaniola, le Mexique, le Paraguay, le Pérou et l'Uruguay. En effet, l'espèce y est bien présente, mais la proximité avec sa zone d'indigénat ne permet pas d'affirmer son statut (Montagnani et al., 2017). Enfin, elle est présente dans de nombreux autres pays à travers le monde (voir Montagnani et al. (2017) pour une analyse complète de la distribution de l'ambroisie à feuilles d'armoïse à travers le monde).

En Europe, elle est considérée comme officiellement invasive en Autriche, Croatie, République tchèque, France, Portugal, Roumanie, Serbie, Slovaquie, Slovénie, Espagne, Suisse, Ukraine et Royaume-Unis (Essl et al., 2015). Elle est considérée comme naturalisée en Arménie, en Biélorussie (Gusev, 2019), en Bulgarie, en Finlande, en Allemagne (Brandes and Nitzsche, 2006), au Luxembourg, en Moldavie, et en Pologne (Figure 8).

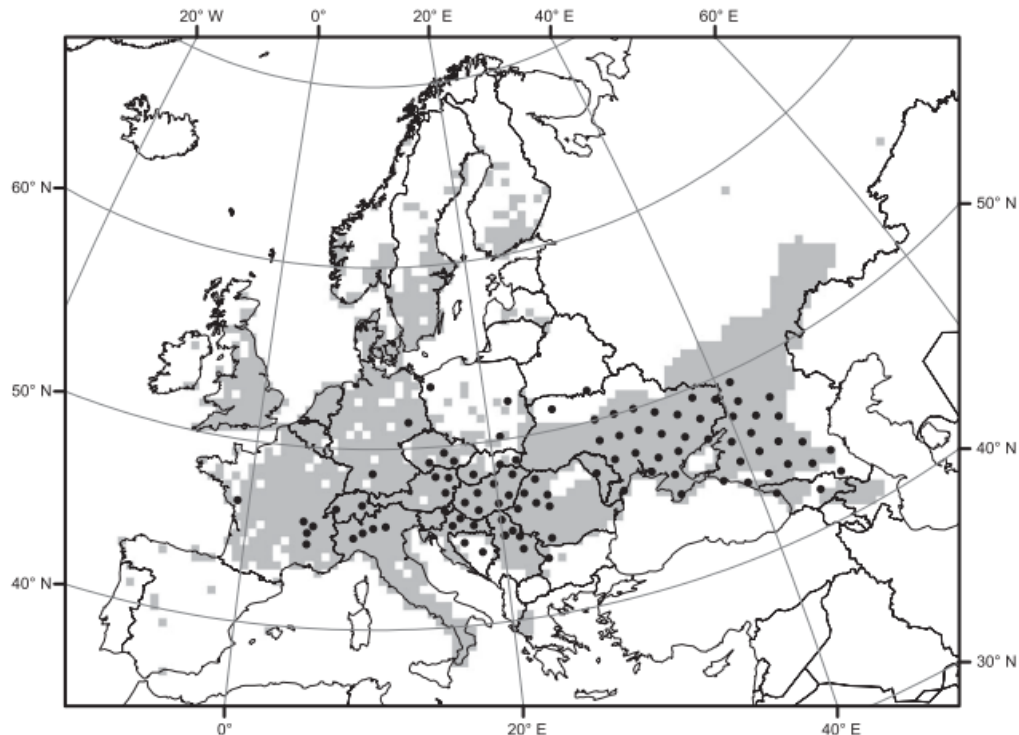


Figure 8 – Distribution d'*Ambrosia artemisiifolia* en Europe (les zones grises de 50 X 50 km sont les zones colonisées). Les points noirs indiquent les zones très touchées par l'espèce, avec un haut niveau d'allergies enregistrées. Carte issue de Essl et al. (2015).

Elle arrive en Europe dès le 19^{ème} siècle (Chauvel et al., 2006; Essl et al., 2015), où elle se répand rapidement et devient une adventice des cultures très importante. En Europe, son invasion a beaucoup été étudiée (Chauvel et al., 2006; Essl et al., 2015, 2009; Kazinczi et al., 2008; Ortmans et al., 2016a; Pinke et al., 2019; Smith et al., 2013) et il est donc possible de distinguer trois grandes zones envahies :

- Le sud de la France (Chauvel et al., 2006; Chauvel and Cadet, 2011; B Fumanal et al., 2008) dans la vallée du Rhône.
- Le nord de l'Italie dans la région du Po (Gentili et al., 2017).
- Une grande zone dans les pays de l'Est : l'ouest de l'Autriche (Essl et al., 2009), le sud de la république tchèque, le sud de la Pologne, la Slovaquie, la Hongrie (Kazinczi et al., 2008; Pinke et al., 2013, 2011) la Slovénie, la Croatie, la Bosnie-Herzégovine, la Serbie, l'Ukraine, et la Roumanie.

Au nord, la plante serait limitée par les faibles températures d'automne qui tuent les plantes matures avant la production de graines ((Chapman et al., 2014; Skálová et al., 2015).

En Belgique, l'espèce est relativement présente en région flamande, et faiblement présente en Wallonie. Cependant, l'espèce est capable d'y établir des populations avec un taux de croissance conséquent (Ortmans et al., 2017) et il est donc fortement conseillé de faire attention aux importations de graines de tournesol (B Fumanal et al., 2008; Genton et al., 2005). Son statut reste cependant incertain, son stade de colonisation n'ayant pas encore été identifié.

Prédictions pour sa future répartition

D'une manière générale, le changement climatique va largement contribuer à la modification des étendues des espèces, favorisant ainsi les EEE, qui se développent plus rapidement dans un nouveau territoire.

Plusieurs études prédisent une expansion de l'espèce en raison du changement climatique (Essl et al., 2009). Les modèles à échelle continentale semblent montrer que les étés plus chauds et les gelées d'automne tardives favoriseront son expansion au nord. Le changement climatique a déjà permis à la plante d'atteindre les îles britanniques et le sud de la Scandinavie (Chapman et al., 2014; Cunze et al., 2013). À l'inverse, l'extrême sud de l'Europe devrait devenir défavorable dû aux inondations d'été, les trop hautes températures (Essl et al., 2015) et les précipitations trop faibles en été (Chapman et al., 2014).

L'espèce est connue pour être favorisée par l'augmentation de la température (Ziska and Caulfield, 2000), ainsi que par l'augmentation de la concentration de dioxyde de carbone. De cette manière, il est attendu que son aire de répartition s'élargisse en Europe, et se déplace vers le nord dû au changement climatique.

Construction d'un modèle extrapolé

Les différentes manipulations seront effectuées sur R (R Core Team, 2020) et QGIS (QGIS.org, 2020).

Données utilisées

Les données de présence utilisées ont été récupérées sur le site Gbif (Global Biodiversity Information Facility ; Gbif.org (2020)) pour les Etats-Unis. Les données de base du modèle sont choisies pour la zone d'indigénat à défaut de données exploitables pour l'Europe. En effet, les données européennes souffrent d'une très forte hétérogénéité dans l'échantillonnage. Bien que le Canada soit aussi considéré comme zone d'indigénat par Basset et Crompton (1975), ce territoire n'a pas été retenu. Il n'existe presque aucune documentation récente sur la répartition de l'ambrosie au Canada à l'exception du Québec. Ainsi, les données canadiennes de Gbif sont très pauvres, et peu représentatives de la distribution réelle de l'ambrosie.

Concernant les données climatiques, elles proviennent du site Worldclim (Fick and Hijmans, 2017), qui possède une base de donnée bioclimatique très riche. Elle comporte 19 variables (Tableau 1), qui ont été récupérées dans leurs versions 2.1 pour la période 1970-2000 délivrée en janvier 2020. La résolution choisie est 2,5 arc minutes, ce qui correspond à approximativement 5 km².

Tableau 1 – Signification des différentes variables bioclimatiques issues de Worldclim 2.1

Bio1	Annual Mean Temperature
Bio2	Mean Diurnal Range (Mean of monthly (max temp - min temp))
Bio3	Isothermality (BIO2/BIO7) (* 100)
Bio4	Temperature Seasonality (standard deviation *100)
Bio5	Max Temperature of Warmest Month
Bio6	Min Temperature of Coldest Month
Bio7	Temperature Annual Range (BIO5-BIO6)
Bio8	Mean Temperature of Wettest Quarter
Bio9	Mean Temperature of Driest Quarter
Bio10	Mean Temperature of Warmest Quarter
Bio11	Mean Temperature of Coldest Quarter
Bio12	Annual Precipitation
Bio13	Precipitation of Wettest Month
Bio14	Precipitation of Driest Month
Bio15	Precipitation Seasonality (Coefficient of Variation)
Bio16	Precipitation of Wettest Quarter
Bio17	Precipitation of Driest Quarter
Bio18	Precipitation of Warmest Quarter
Bio19	Precipitation of Coldest Quarter

Pour les extrapolations temporelles, les données ont aussi été récupérées sur Worldclim. Celles-ci ont été préparées dans le cadre du CMIP6 (Coupled Model Intercomparison Project Phase 6) en vue du sixième rapport du GIEC, avec la version 2.1 de Worldclim comme base. Afin d'avoir des données climatiques les plus fiables possibles, elles ont été récupérées pour quatre GCMs (Global Climate Models ; Pierce et al.(2009)) à savoir BCC-CSM2-MR, CAnESM5, CNRM-CM6-1 et MIROC6. La résolution de 2,5 arc minutes et le choix des GCMs ont été conditionnés par la disponibilité des données sur la plateforme Worldclim.

Méthode

Traitement des données

Les données d'occurrence sont traitées avant la modélisation. Premièrement, un premier tri est effectué pour supprimer les points de présence erronés. De cette manière, les points en dehors du territoire des Etats-Unis sont supprimés. Ensuite, les points situés dans les états d'Alaska et à d'Hawaï sont éliminés afin d'améliorer la lisibilité des cartes. La prochaine étape est de supprimer les données présentes en double. En effet, les données Gbif sont parfois récoltées à partir de plusieurs jeux de données, ce qui résulte en un dédoublement d'information.

Afin de réduire les biais d'échantillonnage potentiels, un filtrage spatial est effectué sur le jeu de données. Grâce au package '*spThin*' (Aiello-Lammens et al., 2015), les données de présence sont filtrées pour espacer les données de 10 kilomètres, ce qui correspond à deux cellules. En effet, les points d'occurrence étant très concentrés à l'est du territoire, cela permet de donner une moindre importance aux zones très denses, tout en intégrant cette densité dans le modèle. L'ambrosie étant indigène, les biais d'échantillonnage sont supposés plus réduits.

Les données bioclimatiques mondiales sont découpées pour les Etats-Unis afin de les utiliser pour les modélisations. Un autre découpage est réalisé pour l'Europe et pour la Wallonie, en vue de l'extrapolation spatiale. Enfin, les données climatiques futures sont fusionnées par GCM afin de n'obtenir que 19 données raster par scénario climatique, puis découpées pour la Wallonie. Afin de garder le même système de coordonnées dans toutes les manipulations, toutes les données sont converties en WGS84.

Génération du background et partitionnement des données

Etant donné l'incertitude quant au degré d'échantillonnage, un nombre de points de background d'environ dix fois supérieur à celui des présences filtrées sera choisi. Ces points seront dans la mesure du possible aléatoirement répartis dans une zone jugée plus défavorable sur base des points de présence. De cette manière, l'algorithme Bioclim (Nix and Busby, 1986) sera lancé avec tous les prédicteurs et l'ensemble des présences, afin de sommairement séparer le territoire des Etats-Unis en deux. L'algorithme Bioclim ne produit pas une classification suffisamment pertinente pour un modèle de prédiction, mais permet ainsi de générer les points de pseudo-absence dans une zone plus défavorable, limitant de cette façon les fausses absences. Enfin, la zone défavorable issue de Bioclim sera amputée d'un tampon de 10 km autour des points de présence qui sont en son sein. Les points seront générés aléatoirement d'une manière stratifiée, c'est-à-dire en évitant de générer deux points dans la même cellule des rasters prédicteurs. Cette méthode en deux étapes sera faite sur R par le package '*dismo*' (Hijmans et al., 2017).

Le territoire américain sera ensuite divisé en cinq blocs géographiquement indépendants afin de répondre au partitionnement des données pour la future validation croisée. De cette manière, le package '*blockCV*' (Valavi et al., 2019) sera utilisé pour créer une grille. Celui-ci est sera créé pour des carrés de 1000 kilomètres de côté à l'équateur. En effet, le package transforme les kilomètres en degrés en divisant la donnée par 111325, la distance standard d'un degré en mètre à l'équateur. De cette manière, les carrés auront 8.9827 degrés de côté, séparant le territoire de manière à générer des cases ayant un nombre comparable de points. Chaque case sera ensuite numérotée de 1 à 5 systématiquement, créant ainsi cinq groupes de cases indépendantes. La validation croisée utilisera chaque groupe comme groupe de validation lors des cinq modèles générés.

Le groupe 1 sera utilisé comme référence de validation pour l'élaboration du modèle principal. De cette manière, les points de présence et de pseudo-absence présents dans les cases du groupe 1 sont séparés du jeu de données afin de créer le jeu de données de validation du modèle principal.

Sélection des variables

La première étape de la sélection des variables est d'exclure les variables fortement corrélées entre elles. Cette étape se révèle être très importante, les variables bioclimatiques étant connues pour être particulièrement corrélées entre elles. Le jeu de données d'entraînement créé, les valeurs des différents rasters prédicteurs seront extraites pour chaque point de présence et de pseudo-absence. Une matrice de corrélation sera ainsi créée en calculant le coefficient de Pearson entre chaque variable. La matrice de corrélation permettra ainsi de créer des groupes de variables fortement corrélées, desquelles une seule variable devra ressortir. Le seuil du coefficient de Pearson choisi pour sélectionner les variables corrélées sera de 0.7. Pour chaque groupe de variables fortement corrélées, une analyse en composantes principales (ACP) sera effectuée. Ainsi, la variable participant le plus à la première dimension de l'ACP sera sélectionnée. De cette manière, le jeu de données sortant de la première sélection se composera des variables peu corrélées et des variables sélectionnées par les ACPs. Pour chaque sélection de variables pendant l'ACP, la variable choisie pourra être sélectionnée manuellement si l'une d'elles semble plus pertinente en fonction de l'espèce.

A l'issue de cette première sélection, le nombre de variables devrait être considérablement restreint. Une première modélisation Maxent sera effectuée pour identifier le taux de participation au modèle de chaque variable. Les variables participant à plus de 10% seront sélectionnées avec un minimum de quatre variables. Dans le cas où quatre variables ne seraient pas sélectionnées, la variable avec la contribution la plus élevée sera choisie. Ce set de variables devra encore subir une dernière sélection.

Le modèle étant destiné à être extrapolé dans l'espace et dans le temps, il se doit d'être le moins complexe possible. Les études portant sur le sujet préconisant tout de même des modèles pluri-variables, un minimum de trois variables finales seront gardées sur le set sortant de la pré-modélisation. Afin de sélectionner le set de variables réduisant la complexité du modèle au mieux, le critère d'information d'Akaike corrigé pour les petits échantillons (AICc) sera calculé pour différents sets de variables (Warren et al., 2014), au moyen du package 'ENMeval' (Muscarella et al., 2014). Ceux-ci seront créés pour avoir les différents sets de variables possibles avec trois variables minimum. Pour chaque set, une modélisation Maxent sera réalisée, permettant ainsi de calculer le nombre de paramètres du modèle à estimer. Ce facteur, couplé aux données de présence et à la carte prédictive de sortie permettra de calculer l'AICc pour chaque set de variables. Celui avec l'AICc le plus faible sera sélectionné pour être le set de variables final.

Modélisation et validation croisée.

La modélisation principale sera effectuée avec le set de variables sélectionné et les points de présence/pseudo-absence d'entraînement. De ce modèle, la participation des variables ainsi que leurs réponses écologiques seront récupérées. Enfin, un raster des valeurs prédictives sera effectué pour le modèle sur le territoire des Etats-Unis.

Afin d'évaluer le modèle, une validation croisée sera effectuée grâce au partitionnement en blocs. Un modèle sera réalisé en prenant à chaque fois un jeu de données de validation différent, selon le bloc géographique. Ainsi, cinq modèles seront générés, dont la contribution des variables et l'AUC seront récupérés. De la même manière, le seuil utilisé par Maxent sera aussi récupéré pour chaque modèle. La moyenne des AUC et son écart-type seront calculés afin de juger de la validité du modèle, ainsi que la moyenne des seuils, qui sera utilisée pour séparer les rasters de valeurs prédictives en zones

favorables et défavorables à l'ambroisie. Une analyse des taux de participation des variables de chaque modèle sera effectuée, afin de juger de l'homogénéité du modèle dans l'espace.

Grâce à ce seuil, une carte des zones favorables et défavorables à l'ambroisie à feuilles d'armoise aux Etats-Unis sera générée.

Extrapolations spatiales et temporelles

Le modèle généré et le seuil défini, il sera possible d'extrapoler le modèle à un autre territoire. Pour cela, les valeurs prédictives seront calculées pour les mêmes sets de variables, mais découpés à l'Europe dans un premier temps, puis à la Wallonie. De cette manière, il sera possible de comparer les zones favorables identifiées par le modèle en Europe aux zones effectivement envahies et juger ainsi de la transférabilité du modèle en Europe, et donc en Wallonie.

Enfin, les sets de variables changeront pour laisser place à ceux des quatre scénarios climatiques découpés pour la Wallonie, afin d'identifier les futures zones potentiellement favorables à l'installation de l'ambroisie annuelle.

Construction d'un modèle indépendant

Données utilisées

Les données de présence utilisées sont issues de trois sources. D'une part, les occurrences du site Gbif pour la Wallonie, celles du département de l'Etude du milieu naturel et agricole (DEMNA) et enfin celles de la plateforme Observation.be.

Les variables environnementales '*Lifewatch-WB ecotope database*', qui est le jeu de données le plus complet pour la Wallonie à ce jour, seront utilisées comme variables d'entrées potentielles. Ce jeu de données est en libre accès sur <http://maps.elie.ucl.ac.be/lifewatch/ecotopes.html> et comprend 110 variables environnementales avec pour unité l'écotope. Ces 110 variables comprennent les 19 variables bioclimatiques de Worldclim, des variables liées à la neige dérivées de '*Lifewatch-WB snow analysis*', des variables topographiques, des variables liées à la couverture de sol, des variables de paramètres de sol (type, profondeur, drainage), la hauteur dérivée d'un jeu de données LIDAR, la lumière artificielle et la distance à différents objets (routes, rivières, forêts, ...).

Pour plus d'informations sur les variables '*écotope*' de Lifewatch, se référer à http://maps.elie.ucl.ac.be/download/ecotopes_documentation_en_v29.pdf.

Méthode

Bien que les données utilisées pour le modèle indépendant soient très différentes du modèle extrapolé, de nombreux points communs existent entre les deux méthodes utilisées. Pour cette raison, la méthode exposée ici sera faite en comparaison avec la précédente, dans le but d'exposer les problématiques rencontrées lors d'un tel modèle.

Traitement des données

La première étape est de fusionner les trois jeux de données en un seul, dans le même système de coordonnées, qui sera WGS84. Concernant ce jeu de données, les mêmes traitements sont opérés que pour le modèle américain. Ainsi, les points en dehors de la Wallonie ou en double sont supprimés. Le filtrage spatial se fait de la même manière, avec un rayon de trois kilomètres. Ce rayon a été choisi d'une part pour réduire le biais d'échantillonnage au maximum, étant donné que certaines régions étaient très denses en présences, et pour correspondre à l'étendue étudiée. Trois kilomètres représentent donc un rayon suffisant pour réduire la densité de certaines régions, tout en gardant suffisamment de données pour la modélisation.

Les données environnementales étant sous forme de vecteurs, une rastérisation sera effectuée avec le programme ArcGIS avec pour consigne une résolution de 500 mètres, en prenant pour valeur la

moyenne des données situées dans chaque cellule. Chaque variable sera ainsi transformée en raster utilisable pour la sélection de variables et la modélisation. Ces variables seront ensuite mises dans le même système de coordonnées que les points d'occurrence, à savoir WGS84.

Génération du background

La situation de la Wallonie est complexe quant à la distribution de l'ambrosie à feuilles d'armoise. La situation n'étant pas à l'équilibre, dans des proportions qui nous sont encore inconnues, il est préférable de réaliser un background totalement aléatoire afin d'éviter au maximum les fausses absences. De la même manière que pour le modèle extrapolé, une zone tampon de trois kilomètres sera néanmoins créée autour des points de présence, bien que certains d'entre eux soient de potentielles fausses présences. Environ dix fois plus de pseudo-absences que de points de présence seront générés, toujours d'une manière stratifiée.

Le territoire sera divisé en une grille de 0.4491 degrés de côté, ce qui correspond à 50 km à l'équateur, grâce au package '*blockcv*'. Chaque case porte un numéro systématiquement choisi de 1 à 5, créant ainsi cinq groupes géographiquement indépendants. Cette méthode sera moins efficace étant donné le faible nombre d'occurrences et leur répartition, mais reste un bon moyen de partitionner le territoire. Le groupe 1 sera utilisé comme référence pour le jeu de données de validation, les autres groupes seront fusionnés pour créer le jeu de données d'entraînement pour le modèle principal.

Sélection des variables

Dans un premier temps, les variables issues de Lifewatch seront triées manuellement selon leur pertinence. Les variables non continues ou catégorielles sont exclues, ainsi que les variables considérées comme non pertinentes ou redondantes.

La suite de la sélection de variable suit la même direction que pour le modèle extrapolé. Les valeurs des variables sont extraites pour chaque point de présence et pseudo-absence, puis une matrice de corrélation est créée pour sélectionner les variables les moins corrélées et identifier les groupes de variables fortement corrélées entre elles. Le nombre de variable étant très grand, l'identification de ces groupes peut se révéler difficile. Pour plus de facilité, la matrice de corrélation sera analysée ligne par ligne. Pour chaque ligne, une ACP sera faite pour les variables avec un coefficient de Pearson supérieur à 0.7 et la variable participant le plus à la première dimension est retenue, tandis que les autres sont supprimées. Une nouvelle matrice de corrélation est créée, jusqu'à ce que toutes les lignes de la matrice soient analysées et qu'il ne reste que les variables peu corrélées.

Malgré cette deuxième sélection, le nombre de variables sera encore trop nombreux. Afin d'avoir une meilleure lisibilité, une première pré-modélisation sera créée afin d'identifier les variables contribuant à plus de 5% au modèle, qui seront préservées, les autres étant considérées comme peu utiles au modèle. Afin de réduire le nombre de variables, une seconde pré-modélisation sera effectuée avec les variables restantes pour ne garder que les variables participant à plus de 10%. Tout comme pour le modèle extrapolé, un minimum de quatre variables sera retenu. Étant donné que le modèle est indépendant, il est moins important qu'il soit moins complexe. En effet, il n'a pas besoin d'être extrapolé. Cependant, les modèles moins complexes ont tendance à être plus fiables, et facilitent l'interprétation, ce qui explique le choix pris ici de réduire le nombre de variables.

Dans la même optique, l'AICc sera aussi utilisé pour choisir le set de variables final. Les sets de variables seront créés pour avoir les différents sets possibles, avec trois variables minimum. Après une modélisation Maxent sur chaque set, le nombre de paramètre puis l'AICc sera calculé. Étant donné la haute complexité supposée des modèles et le faible nombre d'occurrences, il est possible que le calcul de l'AICc soit impossible ; un choix selon le nombre de paramètres sera donc effectué dans ce cas.

Modélisation

La modélisation principale sera effectuée avec le set de variables sélectionné et les points de présence/pseudo-absence d'entraînement. Tout comme pour le modèle extrapolé, la contribution des variables et leurs réponses écologiques seront récupérées, et un raster des valeurs prédictives sera créé pour la Wallonie.

La validation croisée sera faite de la même manière que pour le modèle précédent. Les cinq modèles générés produiront ainsi un graphique des contributions des variables, une évaluation grâce à l'AUC et un seuil de distinction des présences. L'AUC moyen et son écart-type seront calculés, ainsi que le seuil moyen. Celui-ci sera utilisé pour générer une carte des zones favorables et défavorables à l'ambrosie à feuilles d'armoise en Wallonie.

Etant donné la situation de non-équilibre et les biais d'échantillonnage importants, il est attendu que la validation du modèle soit peu efficace.

Résultats

Modèle extrapolé

Filtration spatiale et sélection des variables

A l'issue de la filtration spatiale, le nombre de présences d'*Ambrosia artemisiifolia* aux Etats-Unis est de 1213. Le nombre de points de background généré sera donc de 10 000, ce qui correspond à un ratio de présences/pseudo-absences d'environ 1 :10.

Concernant la sélection de variables pour le modèle, la matrice de corrélation a éliminé douze variables fortement corrélées avec un coefficient de Pearson inférieur à 0,7. Les variables participant à plus de 10% au modèle préliminaire sont bio2, bio5, bio15 et bio18 (figure 9 ; Tableau 2). Etant donné que seules quatre variables sont sélectionnées, aucune sélection supplémentaire n'est effectuée avant la sélection selon la complexité.

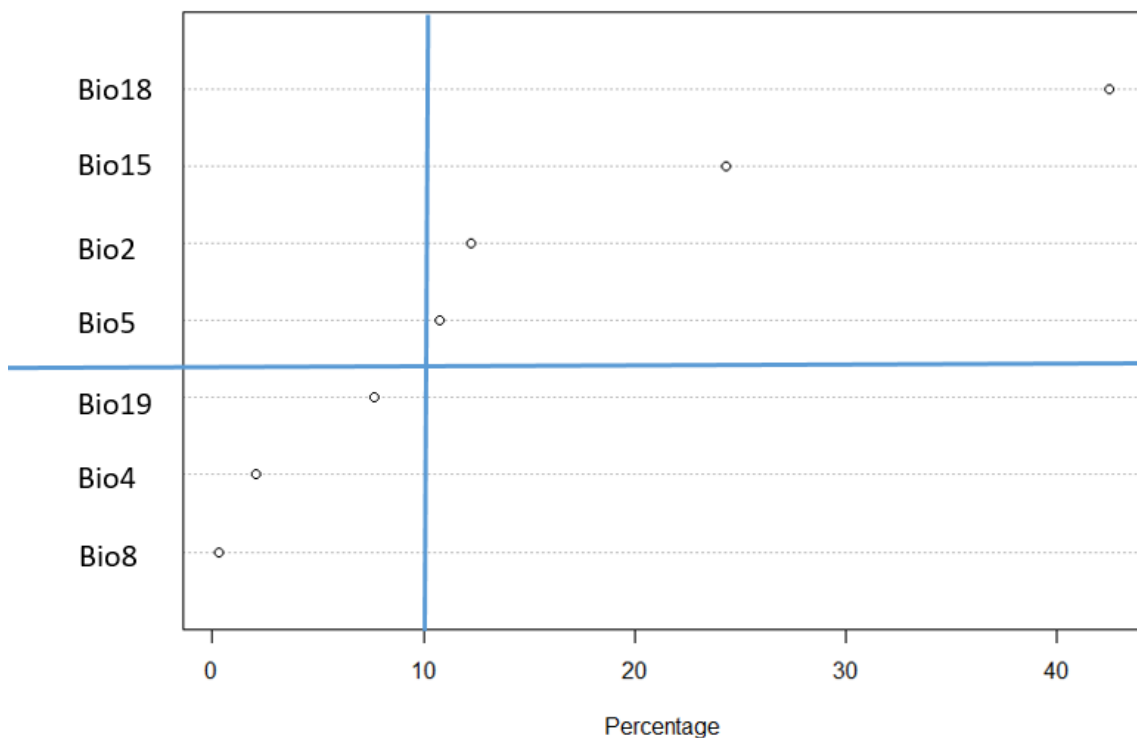


Figure 9 – Pourcentages de participation au modèle préliminaire en vue de sélectionner les variables nécessaires au modèle à extrapoler. Les variables sont en ordonnée avec bio18 les précipitations du quartier le plus chaud de l'année, bio15 le coefficient de variation des précipitations, bio2 la moyenne des amplitudes des températures mensuelles, bio5 la température maximale du mois le plus chaud, bio19 les précipitations du quartier le plus froid de l'année, bio4 l'écart-type des températures et bio8 la température moyenne du quartier le plus humide de l'année. En bleu les lignes qui séparent les variables retenues et exclues.

Tableau 2 – Signification des quatre variables bioclimatiques retenues à l'issue de la sélection par participation au modèle préliminaire.

Bio2	Mean Diurnal Range (Mean of monthly (max temp - min temp))
Bio5	Max Temperature of Warmest Month
Bio15	Precipitation Seasonality (Coefficient of Variation)
Bio18	Precipitation of Warmest Quarter

Le nombre de paramètres et la valeur de l'indice AICc pour chaque set de variables généré avec ces quatre variables sont regroupés dans le tableau 3. Le nombre de paramètres le plus faible est attribué au set de variables (Bio5 x Bio15 x Bio18). Ce dernier minimise aussi l'AICc avec une valeur de 31147.87, et est donc sélectionné pour participer à la modélisation finale.

Tableau 3 – Nombre de paramètres et valeurs de l'indice AICc pour chaque set de variable.

Set de variable	Nombre de paramètres	AICc
Bio2 x Bio5 x Bio15 x Bio18	88	31163.73
Bio2 x Bio5 x Bio15	111	31198.13
Bio2 x Bio15 x Bio18	91	31196.63
Bio2 x Bio5 x Bio18	75	31169.02
Bio5 x Bio15 x Bio18	74	31147.87

Modèle

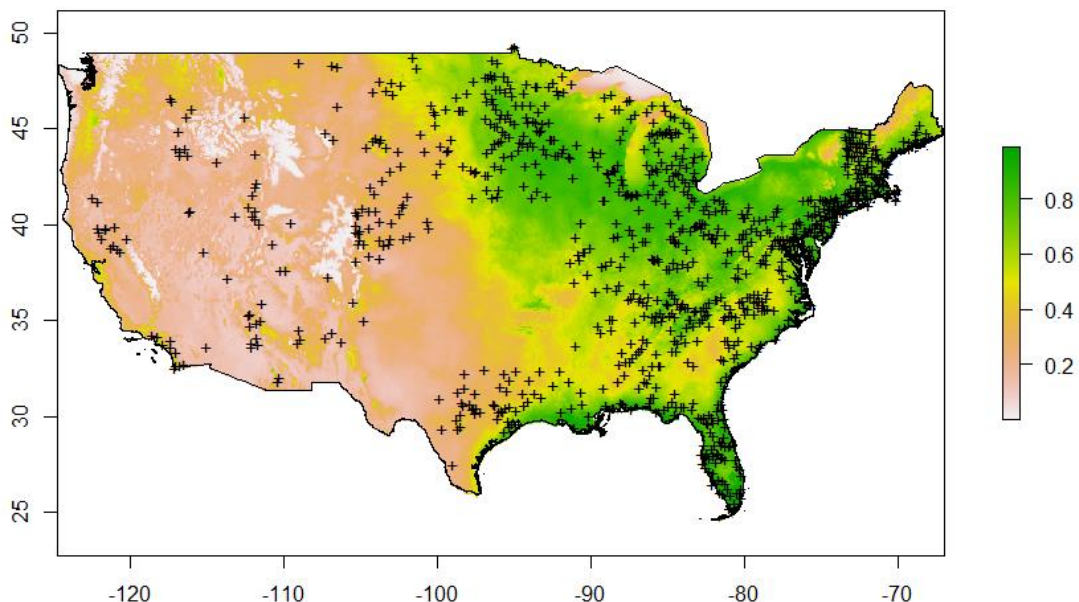


Figure 10 – Carte des valeurs prédictives pour le territoire des Etats-Unis. Les '+' représentent les points de présence.

La carte des valeurs prédictives indique les zones les plus favorables à l'ambroisie annuelle sur la partie est du territoire américain (figure 10). Plusieurs régions à haute densité n'ont pas été retenues par le modèle, comme le Texas ou la côte de Californie. Le seuil déterminé par la validation croisée permettra de scinder la carte en zones favorables et défavorables.

Les réponses écologiques des trois variables retenues sont représentées dans les graphiques de la figure 11, fournis par Maxent.

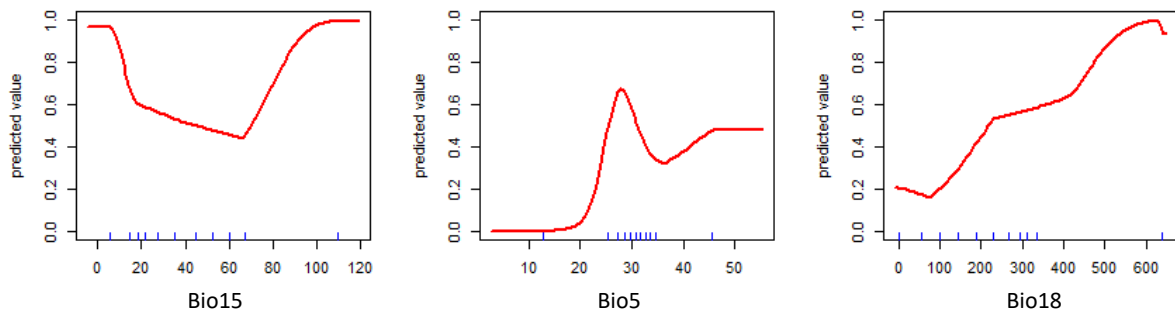


Figure 11 – Réponses écologiques des trois variables retenues pour le modèle à extrapoler. En abscisse les valeurs associées aux variables et en ordonnée la valeur prédictive calculée par le modèle. Les variables bio15, bio5 et bio18 sont respectivement le coefficient de variation des précipitations, la température maximale du mois le plus chaud et les précipitations du quartier le plus chaud de l’année.

Pour chaque variable, la valeur de celle-ci est confrontée à la valeur prédictive du raster de sortie. Plus la valeur prédictive se rapproche de 1, plus l’espèce a de chance d’être présente. Le seuil distinguant les zones favorables et défavorables sera calculé lors de la validation croisée. Cependant, la variable bio15 (coefficient de variation des précipitations) montre une préférence des zones favorables pour les valeurs extrêmes. La variable bio5 (température maximale du mois le plus chaud) grimpe à partir d’environ 20°C, tandis que la variable bio18 (précipitations du quartier le plus chaud) suit une courbe relativement linéaire.

La variable la plus contributive est bio18 (précipitations du quartier le plus chaud de l’année) avec 58.3% de participation, suivi de bio15 (coefficient de variation des précipitations) à 26.3% et bio5 (Température maximum du mois le plus chaud) à 15.5%.

Validation croisée

La séparation du jeu de données en cinq blocs géographiquement indépendants a résulté en cinq modèles différents, permettant une confrontation des différences géographiques du modèle. La carte des Etats-Unis séparée en cinq groupes de blocs de 8,9827° de côté est représentée dans la figure 12.

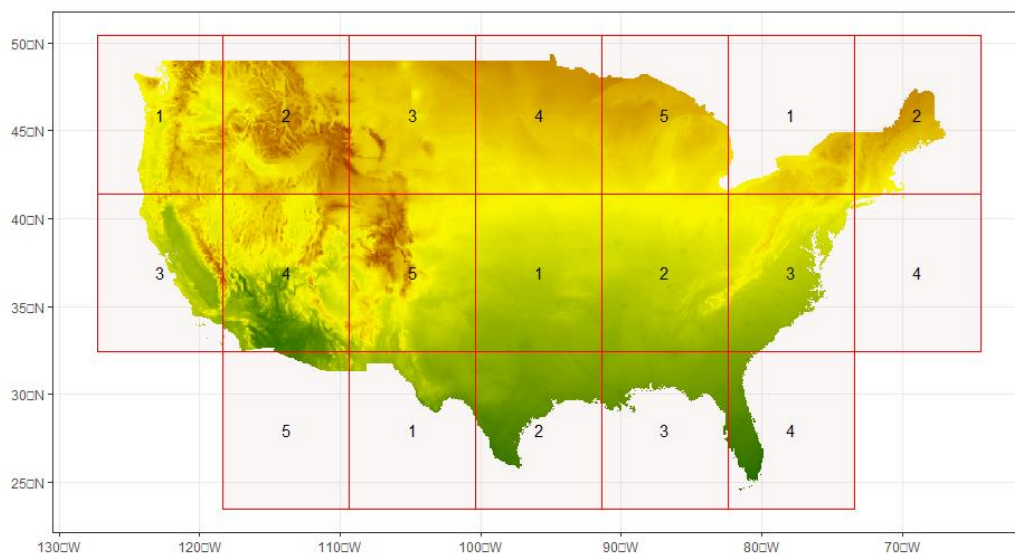


Figure 12 – Visualisation de la séparation en bloc du territoire des Etats-Unis. Le raster d’illustration utilisé est la température moyenne annuelle.

Pour chaque modèle généré pour la validation croisée, les pourcentages de contribution des variables sont illustrés à la figure 13. La variable bio18 reste la plus contributive peu importe le modèle, participant toujours à plus de 50%. Concernant les deux autres variables, la participation est plus hétérogène, avec cependant une participation inférieure à 20% pour bio5. De cette manière, les précipitations du mois le plus chaud est la variable qui contraint le plus le modèle, peu importe la zone géographique.

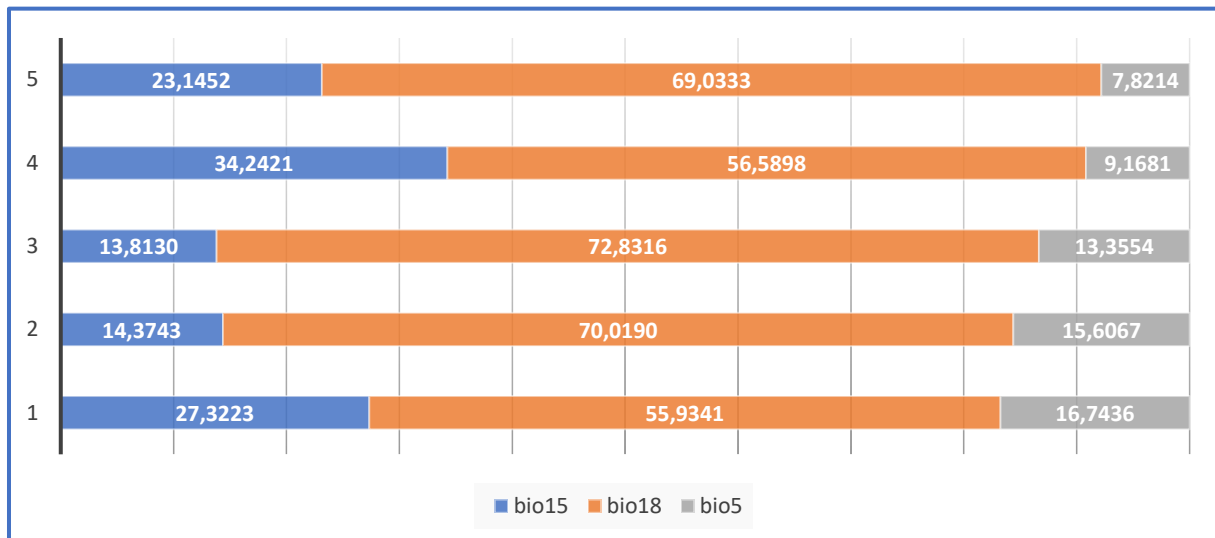


Figure 13 – Participation des variables au modèle final à extrapoler lors de la validation croisée. Chaque groupe est représenté à la verticale par un numéro. La variable bio15 est le coefficient de variation des précipitations et est représentée en bleu ; la variable bio18 est les précipitations du quartier le plus chaud de l'année et est représentée en orange ; la variable bio5 est la température maximale du mois le plus chaud et est représentée en grise.

Les différents AUC calculés oscillent entre 0.695 et 0.83, démontrant une forte hétérogénéité dans la valeur prédictive du modèle (figure 14). La moyenne des AUC calculés est de 0.7692, avec un écart-type de 0.05212. Les seuils calculés à chaque modélisation ont permis d'obtenir un seuil moyen qui sera utilisé pour délimiter les zones favorables et non-favorables. Ce seuil a une valeur de 0.38.

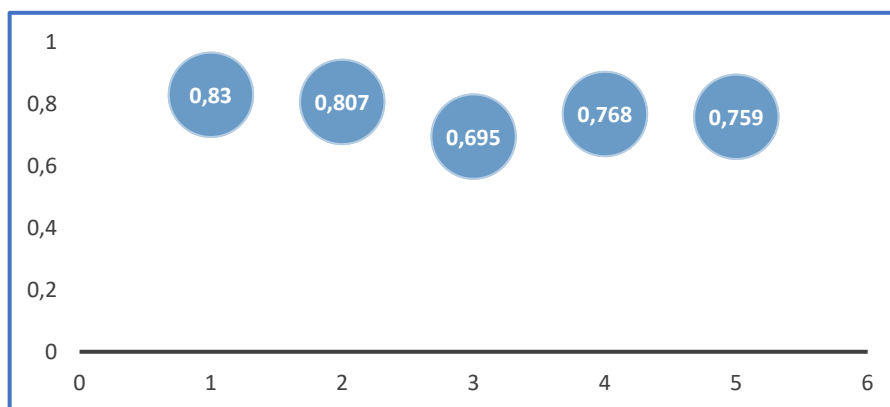


Figure 14 – Valeurs de l'indice AUC pour les 5 modèles de la validation croisée du modèle à extrapoler. Les numéros des groupes sont en abscisses et la valeur d'AUC en ordonnée.

De cette manière, la carte des zones favorables et non favorables pour les Etats-Unis est illustrée dans la figure 15. Comme attendu, la région Est est considérée comme favorable. Quelques zones au nord-ouest, sur la côte californienne et dans les environs du Nouveau-Mexique.

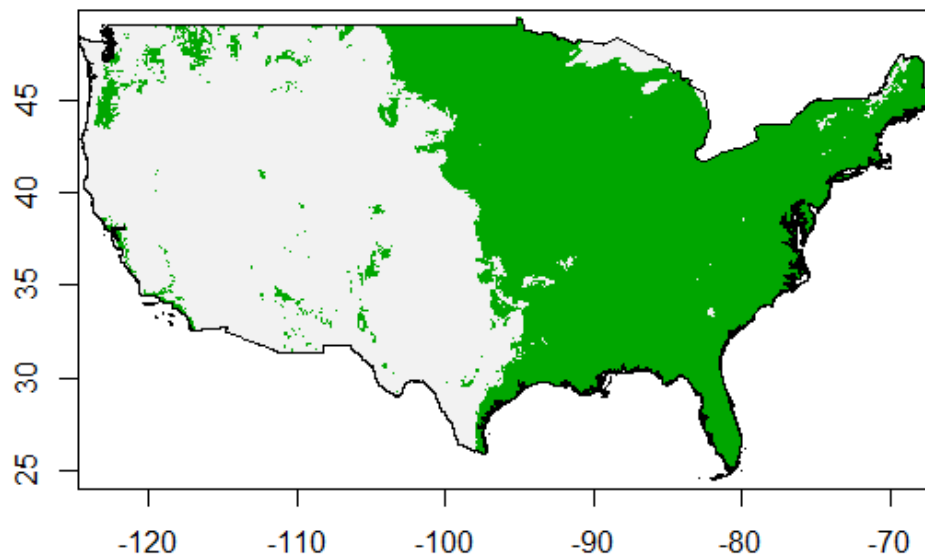


Figure 15 – Carte des zones favorables et défavorables à l’ambroisie à feuilles d’armoise aux Etats-Unis, sur base des valeurs prédictives calculées par le modèle à extrapoler. Le seuil choisi pour séparer les deux zones est de 0.38. En vert la zone favorable, et en gris la zone défavorable.

Extrapolations spatiales et temporelles

A partir de ce modèle et de ce seuil, les zones favorables à l’ambroisie ont été projetées pour l’Europe (figure 16) et la Wallonie (figure 18). Plusieurs régions européennes sont considérées favorables par le modèle, à savoir le centre et le sud de la France, le nord de l’Italie et plusieurs pays de l’Est (la Serbie, la Croatie, la Bulgarie, la Bosnie-Herzégovine, l’Autriche et la Slovaquie particulièrement).

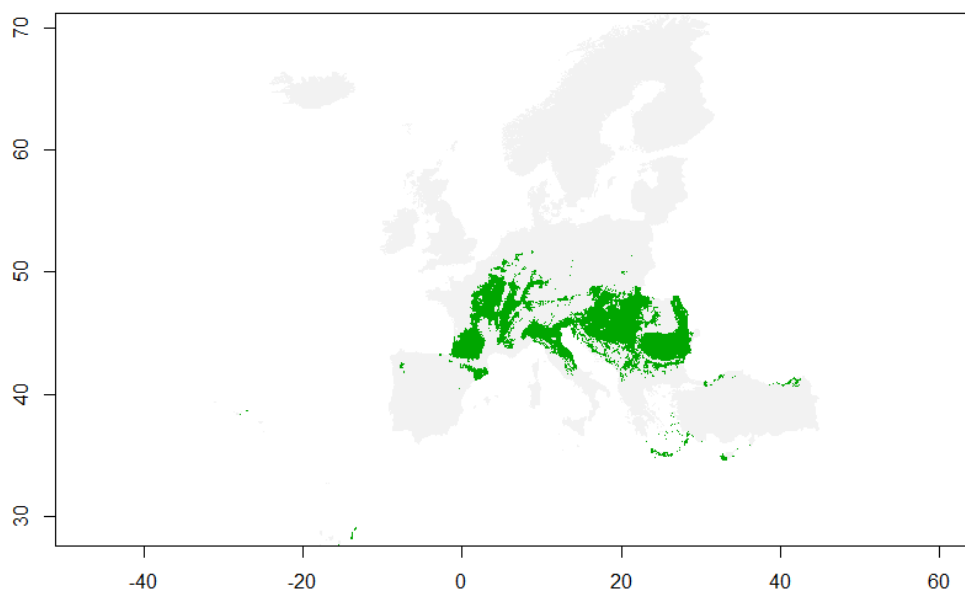


Figure 16 - Carte des zones favorables et défavorables à l’ambroisie à feuilles d’armoise en Europe, sur base des valeurs prédictives calculées par le modèle à extrapoler. Le seuil choisi pour séparer les deux zones est de 0.38. Les zones favorables sont représentées en vert et les zones défavorables en gris.

La carte des valeurs prédictives de la région wallonne indique une vaste zone proche du seuil incluant les provinces de Namur, du Brabant Wallon, le Hainaut (à l'exception du nord-ouest), l'ouest des provinces de Liège et du Luxembourg (figure 17).

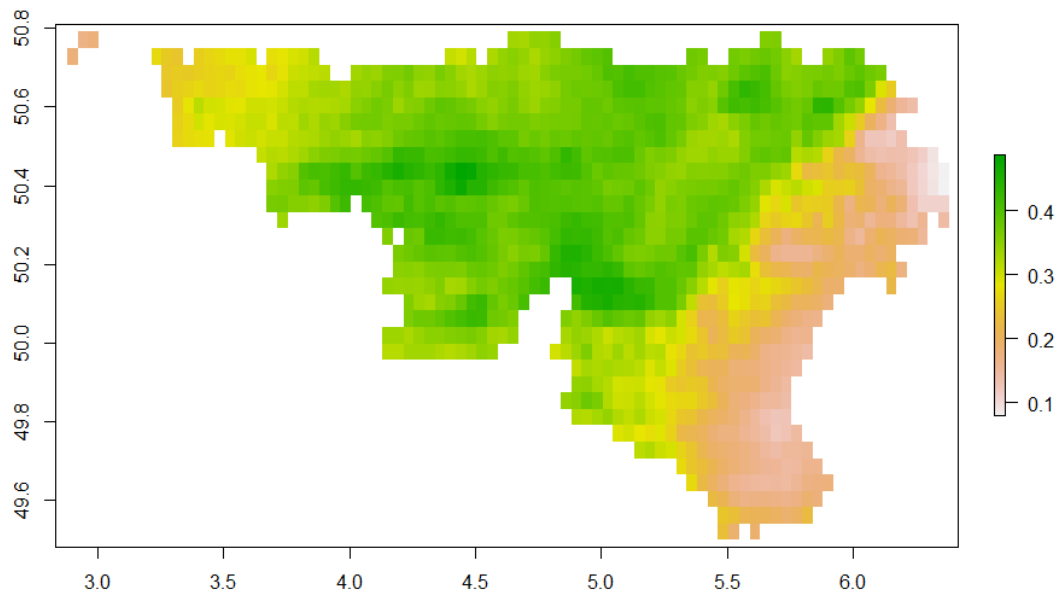


Figure 17 - Carte des valeurs prédictives pour la région wallonne pour le modèle extrapolé.

Concernant la carte binaire, le modèle a identifié une large région centrale traversant le Hainaut et Namur, ainsi que quelques zones favorables en province de Liège (figure 18).

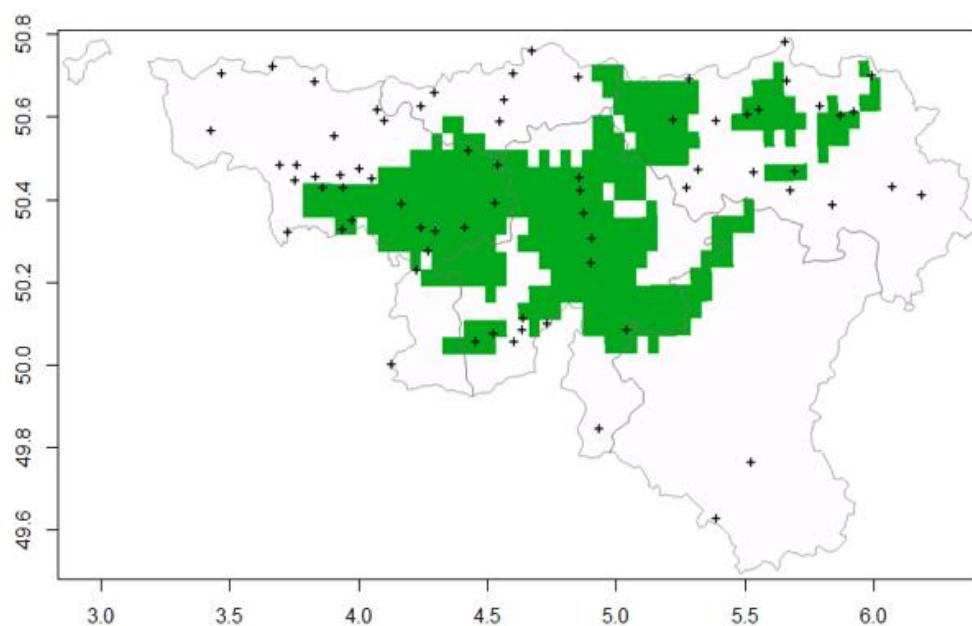


Figure 18 - Carte des zones favorables et défavorables à l'ambroisie à feuilles d'armoise en région wallonne, sur base des valeurs prédictives calculées par le modèle à extrapoler. Le seuil choisi pour séparer les deux zones est de 0.38. Les '+' représentent les points de présence de la région wallonne, non utilisé dans le modèle. Les zones favorables sont représentées en vert et les zones défavorables en blanc.

Suite à ces extrapolations spatiales, les extrapolations temporelles ont été effectuées pour la Wallonie pour chaque scénario climatique (figure 19).

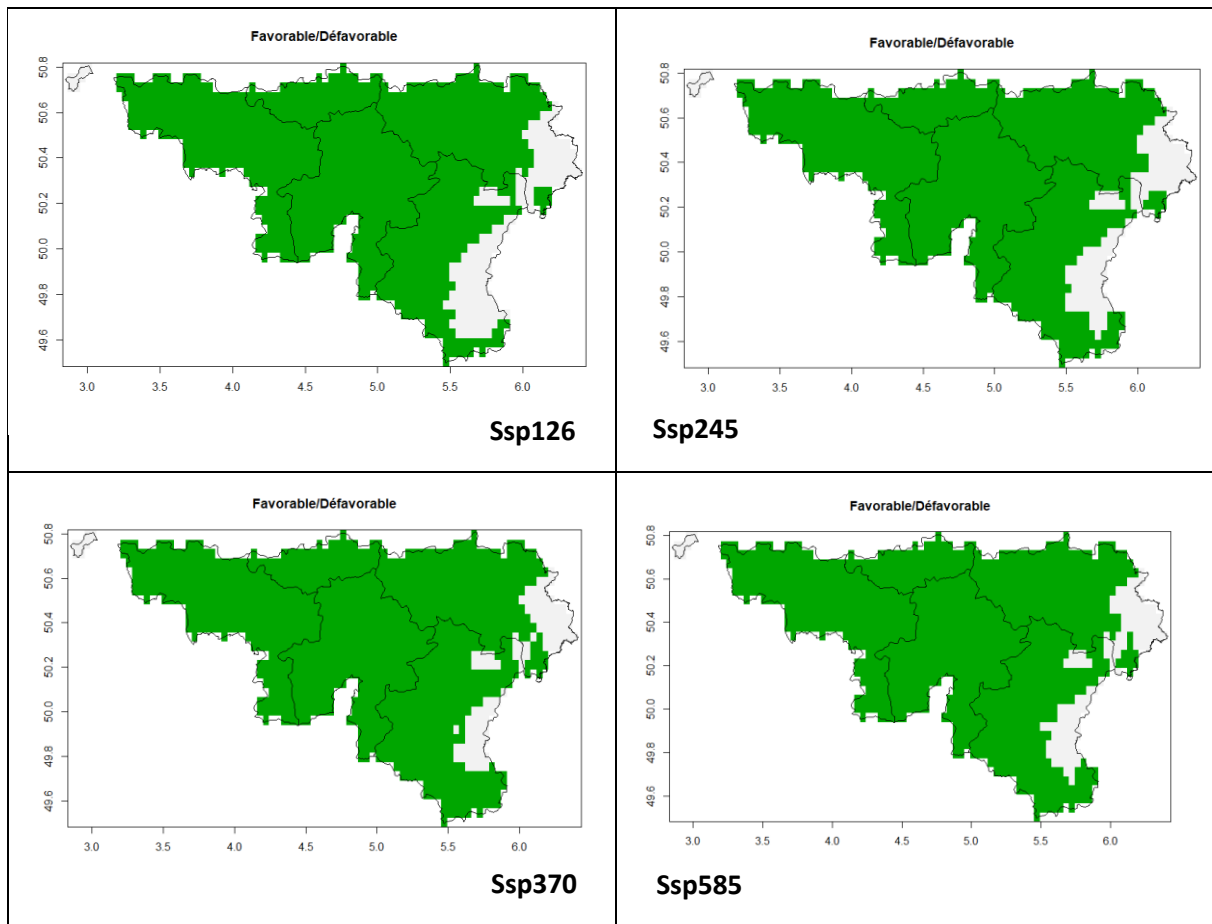


Figure 19 – Cartes des zones favorables et défavorables à l’ambroisie à feuille d’armoise en région wallonne, sur base des valeurs prédictives calculées par le modèle extrapolé dans le temps pour les scénarii climatiques ssp126, ssp245, ssp370 et ssp585 du GIEC. Le seuil choisi pour séparer les deux zones est de 0.38. Les zones favorables sont représentées en vert et les zones défavorables en blanc.

Pour chaque scénario, la presque totalité du territoire devient favorable à l’ambroisie à feuilles d’armoise. Seuls les cantons de l’Est et l’est de la province du Luxembourg semblent rester en zone défavorable, en plus ou moins grande proportion selon le scénario climatique.

Modèle indépendant

Filtration spatiale et sélection des variables

A l'issue de la filtration spatiale, le nombre de présences d'*Ambrosia artemisiifolia* en Wallonie est de 70. Le nombre de points de background généré sera donc de 1000.

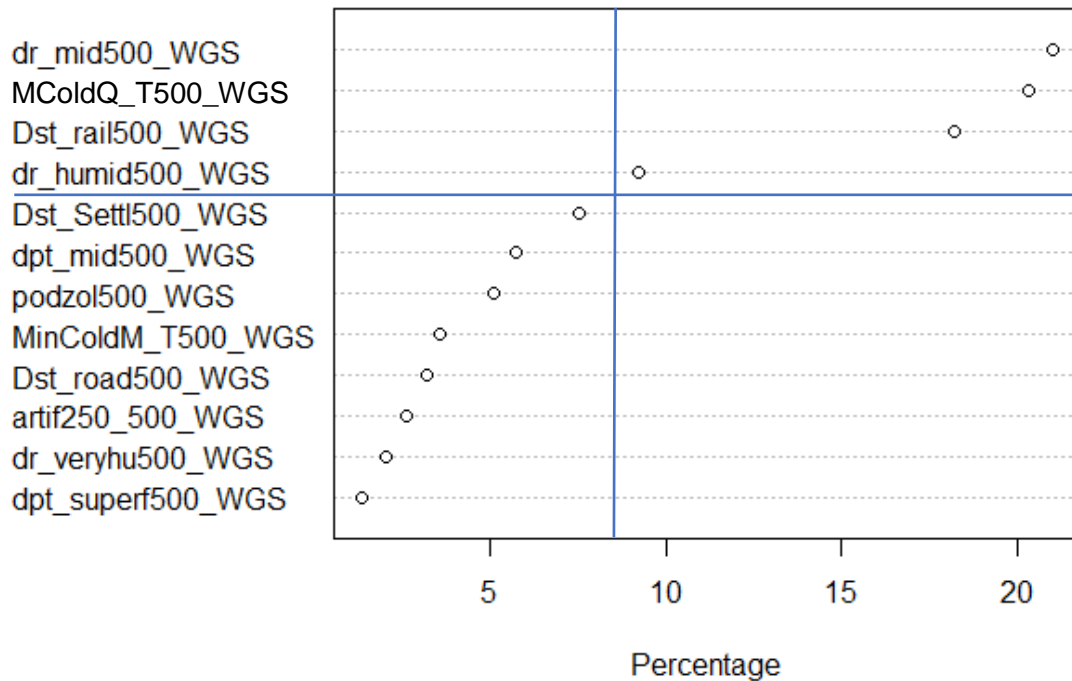


Figure 20 - Pourcentages de participation au second modèle préliminaire en vue de sélectionner les variables nécessaires au modèle indépendant. Les variables sont en ordonnée avec dr_mid500_WGS la proportion de sols à drainage moyen ; MColdQ_T500_WGS la température moyenne du quartier le plus froid de l'année ; Dst_rail500_WGS la distance aux rails de chemin de fer ; dr_humid500_WGS la proportion de sols à drainage humide ; Dst_Settl500_WGS la distance aux habitations ; dpt_mid500_WGS la proportion de sols à profondeur moyenne ; podzols500_WGS la proportion de podzol ; MinColdM_T500_WGS la température minimale du mois le plus froid de l'année ; Dst_road500_WGS la distance aux routes ; artif250_500_WGS les milieux artificiels ; dr_veryhu500_WGS la proportion de sols à drainage très humide ; dpt_superf500_WGS la proportion de sols de faible profondeur. En bleu, les lignes qui séparent les variables retenues et exclues.

Concernant la sélection de variables pour le modèle, la matrice de corrélation a éliminé 14 variables fortement corrélées (coefficient de Pearson > 0.7). Les variables participant à plus de 5% au modèle étaient au nombre de douze. Suite au deuxième modèle préliminaire, les quatre variables retenues sont la distance aux rails de chemin de fer, la température moyenne du quartier le plus froid de l'année, le drainage (moyen et humide) (figure 20). En effet, étant donné que seules trois variables ont participé à plus de 10%, la variable de drainage moyen a aussi été sélectionnée pour avoir un nombre de quatre variables.

Les données de drainage ressortant considérablement dans la sélection de variables, il est pertinent de s'attarder sur la construction de ces variables (Tableau 4). Celles-ci sont les proportions d'une classe de drainage de sol au sein d'un écotope, moyenné pour un carré de 500 mètres de côté.

Tableau 4 – Signification des variables de drainage de jeu de donnée Lifewatch – Ecotope. Issu de la documentation de la version 2.9

Classe	drainage	Definition
Dr_Dry	a ; A (a+b+c+d) & texture=Sandy ; b & texture=Sandy	Very dry « sandy » (cfr 4.1) soils.
Dr_Mid	b & texture≠Sandy ; B & texture≠Sandy ; A & texture≠Sandy ; c	Favourable (clay or silt) or moderate drainage
Dr_Humid	F (e+f) ; h ; f ; e ; l(h+i) ; D(c+d) ; d	Imperfect or poor drainage
Dr_VeryHu	g ; G (e+f+g) ; texture=1/2	Very humid soil (or peatland soil)

Les valeurs de l'AICc pour les cinq sets de variables générés n'ont pas pu être calculés étant donné le trop grand nombre de paramètres. En effet, le calcul de l'AICc nécessite un nombre de paramètres inférieur au nombre de présences (70). Dans ces conditions, le set de variables générant le moins de paramètres a donc été retenu (Tableau 5).

Tableau 5 - Nombre de paramètres et valeurs de l'indice AICc pour chaque set de variable du modèle indépendant.

Set de variable	Nombre de paramètres	AICc
Dst_rail x MColdQ_T x dr_mid x dr_humid	94	NA
Dst_rail x MColdQ_T x dr_mid	74	NA
Dst_rail x MColdQ_T x dr_humid	76	NA
Dst_rail x dr_mid x dr_humid	96	NA
MColdQ_T x dr_mid x dr_humid	92	NA

Le set de variable minimisant le nombre de paramètres étant (Dst_rail x MColdQ_T x dr_mid), il est choisi pour la modélisation finale.

Modèle

La carte de sortie du modèle indique des zones plus favorables en suivant les milieux urbanisés, et par extension le réseau hydrographique (Figure 21). De fait, les hautes valeurs prédictives sont principalement regroupées sur une ligne allant de la région de Mons à Liège, en passant par Namur. De plus, le nord du Hainaut et l'ouest du Brabant Wallon possèdent une valeur prédictive assez élevée. Enfin, certaines régions en Fagne-Famenne, le long de la Meuse, dans les cantons de l'Est et au sud de la province du Luxembourg adoptent des valeurs prédictives élevées.

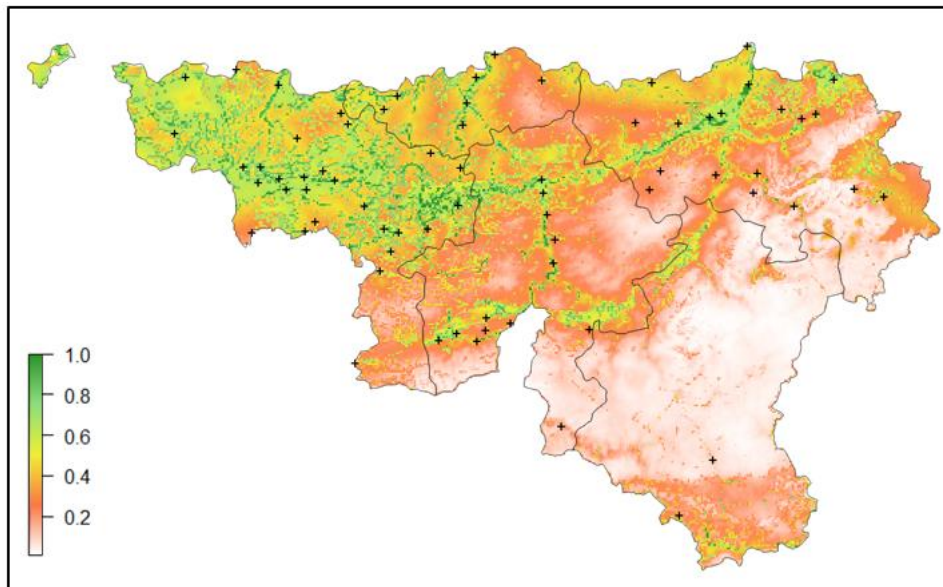


Figure 21 – Carte des valeurs prédictives issues du modèle indépendant pour la région wallonne.

Les réponses écologiques des trois variables retenues sont représentées par les graphiques de la figure 22, fournis par Maxent. Bien que le seuil ne soit pas encore défini, les variables présentent déjà un motif particulier. La valeur prédictive chute rapidement avec l'augmentation de la distance, tandis que la proportion de sols à drainage moyen est la plus favorable à de faibles valeurs, décroissant jusque 40%, puis restant stable. La température moyenne du quartier le plus froid de l'année est un cas particulier, les hautes valeurs correspondant aux erreurs du raster d'origine. Si la lisibilité est faible, il est néanmoins clair que la valeur prédictive augmente avec la température moyenne, qui va de 0 à environ 3.5° Celsius.

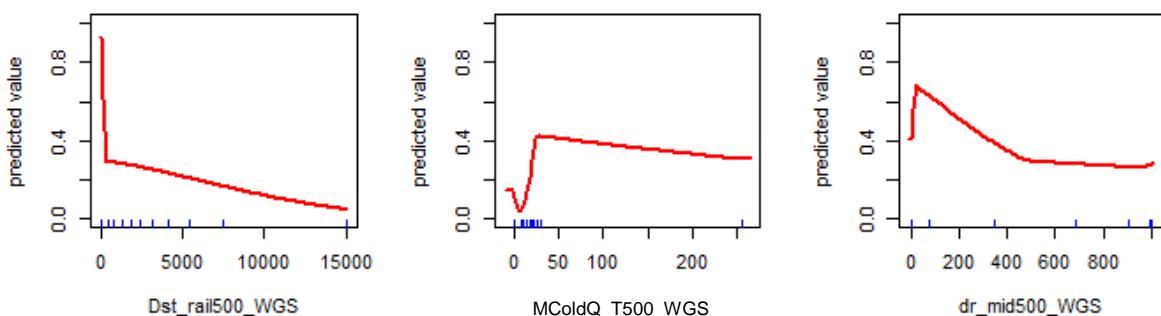


Figure 22 – Réponses écologiques des trois variables retenues pour le modèle indépendant. En abscisse les valeurs associées aux variables et en ordonnée la valeur prédictive calculée par le modèle. Les variables *Dst_rail500_WGS*, *MColdQ_T500_WGS* et *dr_mid500_WGS* sont respectivement la distance aux rails de chemin de fer (en mètres), la température moyenne (en dixième de degré Celsius) du quartier le plus froid de l'année et la proportion de sols à drainage moyen (en %).

La variable la plus contributive est la distance aux rails de chemin de fer, suivi de la température moyenne du quartier le plus froid de l'année. Les deux variables participent respectivement à 40.6% et à 32.6% au modèle. Le drainage participe quant à lui à hauteur de 26.8%.

Validation croisée

La séparation du jeu de données en cinq blocs géographiquement indépendants a résulté en cinq modèles différents, permettant une confrontation des différences géographiques du modèle. Le partitionnement effectué pour la région wallonne en blocs de 0.4491° de côté est illustré dans la figure 23.

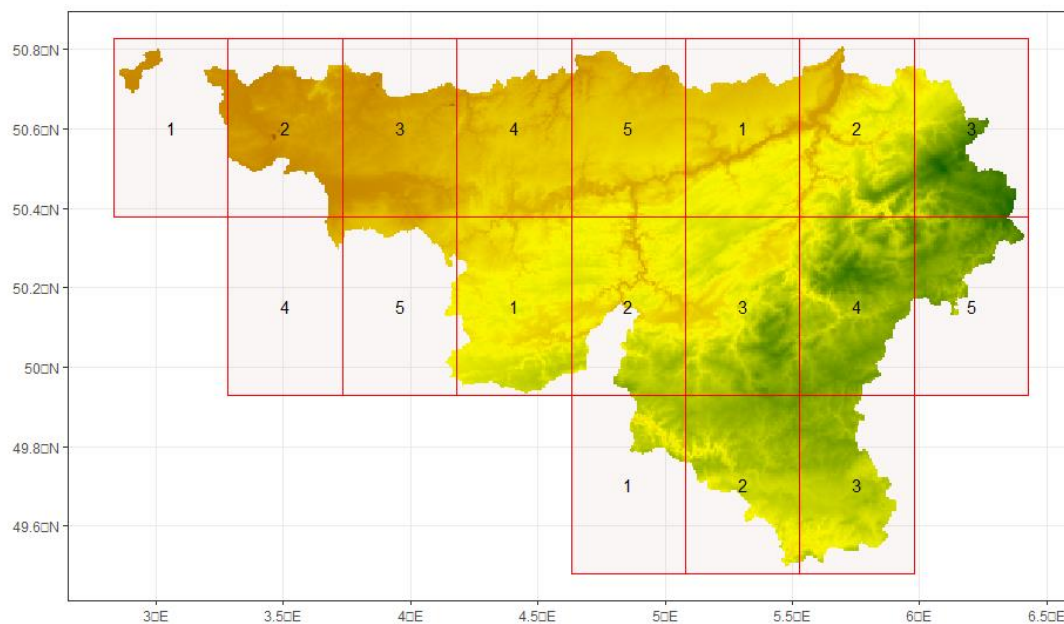


Figure 23 - Visualisation de la séparation en bloc de la région wallonne. Le raster d'illustration utilisé est l'élévation.

Pour chaque modèle généré pour la validation croisée, les pourcentages de contribution des variables sont représentés dans la figure 24. La distribution de la participation est assez hétérogène, la distance aux rails de chemins de fer laissant par deux fois sa place de variable la plus contributive à la température moyenne du quartier le plus froid de l'année. A l'exception du modèle 5, le drainage semble être la variable qui contraint le moins le modèle.

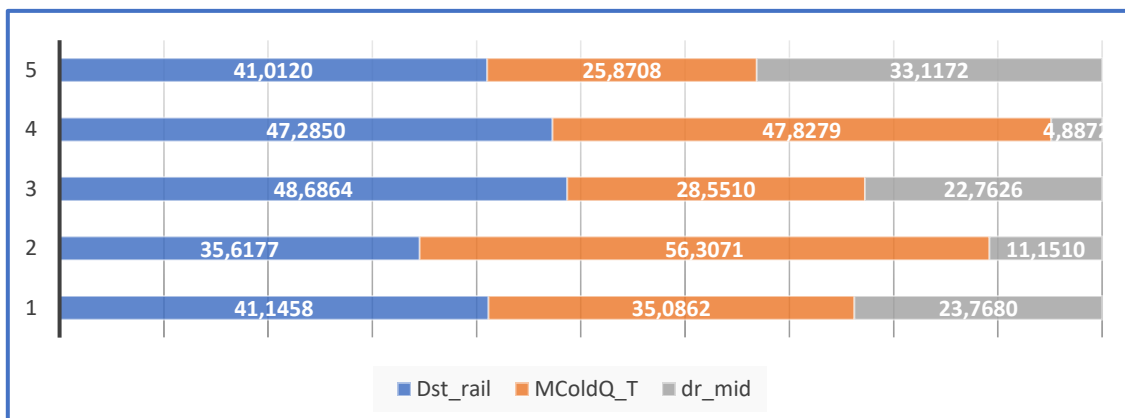


Figure 24 - Participation des variables au modèle final indépendant lors de la validation croisée. Chaque groupe est représenté à la verticale par un numéro. La variable Dst_rail est la distance aux rails de chemins de fer et est représentée en bleu ; la variable MColdQ_T est la température moyenne du quartier le plus froid de l'année et est représentée en orange ; la variable dr_mid est la proportion de sols à drainage moyen et est représentée en grise.

Les différents AUC calculés oscillent entre 0.637 et 0.864, démontrant une très forte hétérogénéité dans la valeur prédictive du modèle (figure 25). La moyenne des AUC est de 0.7330 avec un écart-type de 0.09297. Les seuils calculés à chaque modélisation ont permis d'obtenir un seuil moyen qui sera utilisé pour délimiter les zones favorables et non-favorables. Ce seuil a une valeur de 0.48.

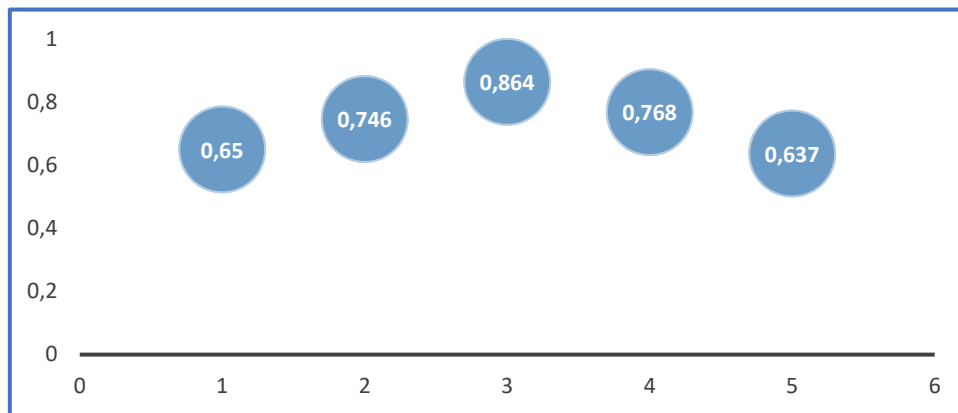


Figure 25 - Valeurs de l'indice AUC pour les 5 modèles de la validation croisée du modèle indépendant. Les numéros des groupes sont en abscisses et la valeur d'AUC en ordonnée.

De cette manière, la carte des zones favorables et non favorables pour la région wallonne est illustrée dans la figure 26. Ces zones suivent, sans surprise, les zones identifiées avec la carte des valeurs prédictives : le long de la Meuse et la Sambre, le nord du Hainaut, l'ouest du Brabant Wallon, les Fagnes-Famenne, le nord de la province de Liège, le nord des cantons de l'Est et le sud de la province du Luxembourg.

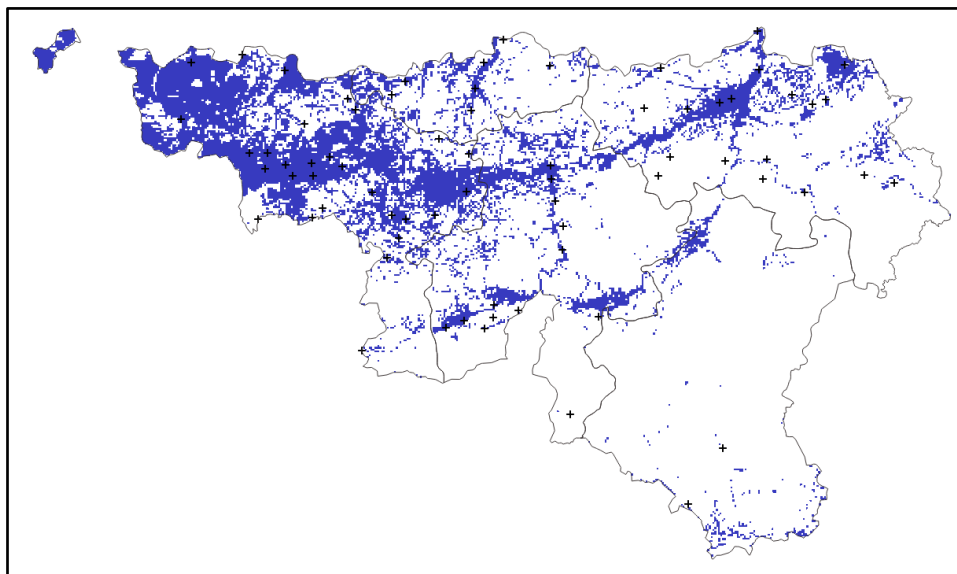


Figure 26 - Carte des zones favorables et défavorables à l'ambrosie à feuilles d'armoise en région wallonne, sur base des valeurs prédictives calculées par le modèle indépendant. Le seuil choisi pour séparer les deux zones est de 0.48. En bleu la zone favorable, et en blanc la zone défavorable.

Discussion

Ambrosia artemisiifolia est une plante très étudiée dans la littérature scientifique, ce qui en fait un exemple de choix pour étudier l'efficacité d'un modèle de par la disponibilité des données la concernant. D'une manière générale, il est attendu que la distribution de l'espèce subira un déplacement vers le nord dû au changement climatique, résultant en une augmentation du territoire potentiellement envahissable.

Afin d'identifier les zones favorables, il existe plusieurs méthodes de modélisation, dont deux ont été traitées lors de cette étude. De telles modélisations présentent de nombreuses limites, liées à la violation de l'hypothèse d'équilibre de l'espèce, aux biais d'échantillonnage et aux fausses données de présence. De cette manière, l'extrapolation d'un modèle créé dans la zone d'indigénat sur la zone d'étude est devenu la norme dans les études de distribution d'espèces exotiques envahissantes. En effet, la zone d'indigénat est souvent considérée comme mieux échantillonnée et plus représentative de la niche de l'espèce. Dans le cas de la Wallonie, une extrapolation de ce modèle signifie une perte de variables potentielles, le territoire wallon étant particulièrement bien fourni en données environnementales. Les deux modèles semblent donc complémentaires, présentant chacun ses qualités et ses défauts.

Il a donc été décidé de réaliser les deux méthodes, en utilisant le territoire des Etats-Unis comme zone d'indigénat pour le modèle extrapolé. Il est intéressant de brièvement discuter du choix de la zone de référence pour le modèle extrapolé. En effet, cette zone aurait pu être l'Europe ou le monde. Ces alternatives permettent entre autres de contrer le phénomène de niche-shift, qui peut se révéler être important pour certaines espèces. Ainsi, prendre un modèle de référence proche géographiquement, tel l'Europe, diminue la différence potentielle de comportement. A l'inverse, prendre comme référence l'intégralité des données disponibles permet de faire une moyenne des niches rencontrées dans le monde entier. Cependant, ces méthodes ne se révèlent efficaces qu'à deux conditions. D'une part, l'espèce étudiée doit subir un niche-shift suffisamment important pour justifier ces manipulations et d'autre part, les données d'occurrence de la zone de référence doivent correctement représenter la distribution de l'espèce. Dans le cas de l'ambrosie à feuilles d'armoise, aucune de ces conditions n'est remplie. En effet, l'espèce ne présente pas ou peu de phénomène de niche-shift (Petitpierre et al., 2012). De plus, le jeu de données Gbif européen est très biaisé concernant *Ambrosia artemisiifolia*. Les données d'occurrence sont en effet très bien renseignées en Allemagne et au nord de l'Europe par exemple, mais très mal dans la majorité des pays de l'Est, pourtant très touchés par l'espèce. Ce type de biais d'échantillonnage conduit à préférer la zone d'indigénat, considérée comme la mieux échantillonnée, comme zone de référence.

Le modèle extrapolé

Le modèle extrapolé consiste en la génération d'un modèle sur le territoire américain, ensuite extrapolé sur le territoire wallon. Le nombre d'occurrences est relativement faible étant donné la taille de la zone d'échantillonnage, mais assez élevé pour être considéré comme robuste. La distribution des données de présence a été considérée comme peu biaisée, mais il est important de noter la grande hétérogénéité de la répartition des points d'occurrence aux USA. En effet, les points de présence sont majoritairement à l'est de la zone d'étude, ce qui peut traduire soit un biais d'échantillonnage, soit une réelle préférence de l'espèce. Dans le cas de l'ambrosie à feuilles d'armoise, la quasi-entièreté du territoire des USA est habituellement considérée comme favorable à *Ambrosia artemisiifolia*, avec une zone d'indigénat historique au centre du territoire (Bassett and Crompton, 1975), bien que certaines

études mentionnent des traces historiques de la présence de l'espèce au sud du Québec, ce qui pourrait étendre la zone d'indigénat historique vers l'est (Lavoie et al., 2007).

Dans cette étude, le faible nombre de présences à l'ouest des USA et la méthode de génération du background a contribué à mettre en lumière les zones optimales plutôt que les zones favorables à l'espèce. De plus, les données d'occurrence canadiennes étant maigres et exclues de l'étude, la résistance aux basses températures qui aurait pu être apportée par les données canadiennes n'a pas pu être correctement incluse, ce qui favorise cette mise en lumière. Cette distinction est un point important à prendre en compte pour la suite de l'étude, l'ambrosie à feuilles d'armoise étant une espèce généraliste assez peu exigeante pour son installation. La distinction effectuée pour les cartes binaires ne traduit donc pas la présence de l'espèce, mais les zones les plus potentiellement touchées. Cette distinction implique qu'une grande partie des points d'occurrence se trouvent dans la zone identifiée '*défavorable*'. Cette étiquette devrait être renommée '*non-optimale*', et cela a grandement impacté la validation du modèle. En effet, la construction de l'indice AUC se base sur la distinction entre des zones favorables et défavorables. Néanmoins, malgré ce défaut, la valeur de l'indice est supérieure à 0.7 et est donc considérée comme bonne selon l'échelle de Elith et al. (2011). L'impact du parti pris d'identifier les zones optimales se ressent néanmoins dans l'écart-type (0.05212) qui traduit donc une forte différence de qualité de prédiction selon la région géographique. Cet écart-type est néanmoins acceptable et l'AUC est satisfaisant pour une espèce s'installant sur plusieurs types de milieux.

Concernant la sélection de variables, il est intéressant de constater que douze variables sur les dix-huit de bases ont été éliminées pour cause d'autocorrélation trop élevée. Cela confirme les nombreuses études mentionnant l'importante autocorrélation des variables issues de Worldclim. Les quatre variables sélectionnées par la contribution au modèle sont la température maximale du mois le plus chaud, les précipitations du quartier le plus chaud de l'année, le coefficient de variation des précipitations et la moyenne de l'amplitude de température mensuelle. Il est possible de catégoriser ces quatre variables en deux groupes : d'une part, la température et les précipitations des mois les plus chauds et d'autre part, la variation des températures et des précipitations. Cela correspond aux résultats de Essl et al. (2009) qui identifient les basses températures moyennes des mois les plus chauds comme le facteur climatique le plus limitant pour l'Europe centrale, et ceux de Chapman et al. (2014) qui expliquent l'absence de l'ambrosie au sud de l'Europe par les faibles précipitations des mois chauds. La contribution au modèle préliminaire tend à donner une plus grande importance aux variables liées aux précipitations. Plus précisément, avec le seuil choisi, le coefficient de variation des précipitations n'intervient pas dans la séparation de la carte des valeurs prédictives en carte binaire. Cependant, le modèle identifie une préférence nette pour les écarts de précipitations extrêmes, défavorisant les zones ayant un coefficient de variation de 20 à 80. Concernant les températures maximales du mois le plus chaud, le seuil définit une température minimale de 25 degrés Celsius pour les zones favorables, ainsi que des précipitations supérieures à 200 mm pour le quartier le plus chaud de l'année. Ces résultats sont logiques quant à la phénologie de l'espèce. Celle-ci peut sortir de dormance en janvier-février, bien que cette période puisse s'allonger jusqu'en mai dans certaines régions envahies (B. Fumanal et al., 2008). La germination a lieu en fin mars-avril, et la floraison en fin juillet-août dans les milieux fortement envahis. De fait, le quartier chaud de l'année joue une importance majeure dans le cycle de vie de l'ambrosie annuelle, ainsi que le mois le plus chaud, qui intervient dans l'induction florale et la production de graines.

La participation au modèle est en adéquation avec les contributions au modèle préliminaire, ce qui tend à montrer une certaine homogénéité dans la contribution des variables. Ce facteur a été vérifié par la validation croisée, qui démontre une participation au modèle relativement homogène, dans le

sens où l'ordre d'importance des variables varie peu. Cependant, les groupes 1 et 5 dénotent des trois autres, par une plus grande participation de bio15 et une plus faible de bio18. La répartition des blocs ne permet pas d'émettre une hypothèse satisfaisante quant à la raison de cette différence.

La carte binaire est conforme aux résultats attendus pour les Etats-Unis, ce qui confirme la validité du modèle. L'extrapolation du modèle sur le territoire européen valide aussi le modèle, étant donné qu'il identifie correctement les zones les plus envahies en Europe. En effet, les zones identifiées sont le centre de la France, la vallée du Rhône et les Pyrénées, le nord de l'Italie ainsi qu'une large zone correspondant aux pays touchés par l'ambrosie à l'est de l'Europe. Ces zones sont particulièrement touchées par *Ambrosia artemisiifolia*, ce qui démontre une bonne extrapolation du modèle, ainsi qu'une évaluation positive. Ces zones identifiées par le modèle ne constituent pas la distribution de l'espèce proprement dite. De cette manière, l'ambrosie est à ce jour présente dans la quasi-totalité de l'Europe. Trois raisons peuvent expliquer ce résultat.

D'une part, la construction du modèle est basée sur l'identification des zones optimales à l'ambrosie, ce qui implique que les zones moins favorables ne sont pas comprises dans la carte binaire, malgré la présence de l'espèce dans ces zones. D'autre part, l'espèce n'est pas en équilibre dans toute l'Europe, contrairement au territoire dont elle est indigène. Ainsi, beaucoup de points d'occurrence sont dans des zones défavorables. Néanmoins, l'espèce a été diagnostiquée invasive dans de nombreux pays d'Europe (Montagnani et al., 2017), ce qui tend à démontrer que l'espèce a atteint sa phase de colonisation voire de limite de capacité et est donc plus en équilibre dans son environnement que lors des premiers stades de colonisation. Enfin, la raison la plus impactante est l'origine des variables elles-mêmes. Les variables bioclimatiques originaires de Worldclim sont construites à partir des données de 1970-2000, ce qui implique un retard de 20 ans sur le climat réel. En 20 ans, le changement climatique a fait son œuvre, ce qui implique que la situation actuelle est intermédiaire entre le modèle référence et le modèle extrapolé dans le temps. Ainsi, les zones optimales devraient avoir migré vers le nord, ce qui concorderait mieux avec la situation actuelle de l'ambrosie en Europe.

Le focus sur la région wallonne révèle une partie du territoire se situant dans l'enveloppe climatique optimale de l'ambrosie à feuilles d'armoise. Ces zones coïncident avec une partie des points d'occurrence récupérés en Wallonie. Seuls les points du nord du Hainaut, du Brabant Wallon et quelques points isolés ne se retrouvent pas dans les zones identifiées. Bien que le nombre de points inclus dans les zones favorables soient faible (seuls 25 points sur 70 sont inclus), une grande partie des points non-inclus sont proches des zones identifiées. L'évaluation d'un modèle par les points d'occurrence du territoire extrapolé n'est pas gage de qualité étant donné la situation ambiguë de l'ambrosie en Wallonie, mais cette correspondance entre les points de présence wallonnes et le modèle est encourageant quant à la qualité prédictive du modèle. Enfin, bien que le seuil choisi sépare la Wallonie en deux zones distinctes, il est pertinent de constater qu'une grande partie du territoire est proche du seuil choisi, ce qui implique une enveloppe climatique proche de l'enveloppe optimale sur la majorité de la Wallonie.

Les résultats de l'extrapolation temporelle sur la région wallonne montrent que la zone optimale à la colonisation de l'ambrosie s'étendra à la quasi-totalité du territoire. L'extrême est de la province de Liège et l'est de la province du Luxembourg restent à l'abri de l'espèce, dans des proportions plus ou moins grandes selon le scénario climatique du GIEC. Ces zones ont de températures estivales plus basses liées à l'altitude, ce qui peut expliquer leur exclusion de la zone favorable.

Le modèle indépendant

Après la filtration spatiale, le nombre de points de présence est de 70. Pour une espèce généraliste, ce nombre est considéré comme faible, un tel nombre ne pouvant représenter correctement la répartition de l'espèce sur le territoire wallon. De plus, la répartition des points de présence est très concentrée, ce qui constitue au mieux une autocorrélation spatiale liée à la colonisation de l'ambrosie, au pire un biais dans l'échantillon. Malgré la faiblesse du jeu de données, plusieurs résultats sortants peuvent être extraits. En effet, le modèle précédent ayant démontré une relative bonne répartition de ces points de présence, les biais liés au modèle seront essentiellement liés au manque de données de présence dû au non-équilibre et/ou au manque d'échantillonnage, ainsi qu'aux fausses absences ou véritables absences en zone favorable liées au stade de colonisation. Les biais liés aux présences sont quant à eux certainement liés au manque d'échantillonnage en zone rurale. En effet, la quasi-totalité des données de présence se trouvent en milieu urbain ou du moins le long de routes ou cours d'eau. Cette présence est logique quant au comportement de l'espèce en début de colonisation, mais constitue un biais quant à sa répartition potentielle. En effet, les sols riches et les milieux ouverts plus distants des milieux urbanisés sont connus pour être favorables à l'espèce, bien que peu touchés en début d'invasion. Une attention particulière devra donc être portée à ces milieux peu échantillonnés et probablement peu touchés vu le stade d'invasion actuel. En effet, même si l'espèce n'a peut-être pas encore atteint ces milieux, son arrivée pourrait être très nocive.

Bien que seules les quatre variables participant le plus au modèle ont été retenues, il peut être intéressant d'analyser les variables retenues pour le second modèle préliminaire. En effet, ces variables ne participent pas suffisamment au modèle pour être sélectionnées, mais ont été retenues parmi les 89 variables pour leur contribution au modèle. En effet, bien que non retenues pour le modèle final, celles-ci sont cohérentes par rapport aux facteurs limitants connus de l'ambrosie à feuilles d'armoise, prouvant une bonne sélection de la part de Maxent malgré la faiblesse du jeu de données.

Premièrement, le drainage semble jouer un rôle important, les trois variables de drainage ayant été retenues dans ces douze variables. Cela confirme le rôle primordial du drainage dans la distribution de l'espèce, identifié par (Gusev, 2019). Ensuite, les variables liées aux milieux urbanisés (milieux artificiels, distances aux routes, distances aux habitations, distances aux rails) sont pertinentes, l'ambrosie annuelle poussant majoritairement sur les sols rudéraux, en bordure de routes ou à proximité des poulaillers dans les régions récemment envahies. D'autre part, ces milieux sont plus visités et ont plus de chance d'être échantillonnés, leur donnant plus de poids dans le modèle. Pour ces raisons, ce sont les milieux anthropisés qui contribuent le plus au modèle. Ainsi, les milieux identifiés par le modèle suivent les milieux urbanisés. Le drainage apporte quant à lui une grande zone au nord du Hainaut, favorisant la présence de l'ambrosie en territoire hennuyer. Le troisième groupe de variable pertinent sont les températures hivernales. En effet, bien que l'espèce passe l'hiver sous forme de graine et soit donc peu touchée par le climat d'hiver, elle a besoin d'une longue période froide pour sortir de sa phase de dormance (Essl et al., 2015; Shrestha et al., 1999). Néanmoins, la réponse écologique fournie par la température moyenne du quartier le plus froid de l'année implique une faible importance de cette variable dans la distribution de l'espèce. Ce résultat est logique, les exigences de l'ambrosie se situant autour de 3.5°C minimum pendant quelques jours à la sortie de l'hiver pour sortir de dormance (Shrestha et al., 1999). En effet, cette situation n'est pas rare en région wallonne, et ne constitue donc pas le facteur limitant le plus la distribution de l'espèce. Enfin, les sols de faible profondeur et les podzols sont les variables les moins pertinentes. L'espèce a en effet peu d'exigences concernant la profondeur de sol. Sa préférence pour les sols plus riches pourrait expliquer la présence de la variable *podzols*, qui agirait négativement sur la présence de l'espèce. Enfin, il est

pertinent de noter qu'aucune variable du premier modèle n'a été retenue par le second, malgré leur pertinence avérée.

Afin de diminuer le nombre de variables en vue de réduire la complexité du modèle, seules les quatre variables participant le plus ont été retenues. De cette manière, le drainage (représenté par deux variables), la température moyenne du quartier le plus froid de l'année et la distance aux rails (qui représente les milieux artificiels) sont les variables qui limitent le modèle final. La dernière étape de la sélection de variable selon la complexité n'a pas pu se faire selon la valeur de l'AICc, le calcul étant impossible pour un nombre de paramètres supérieur au nombre de présences. Avec un nombre de présences de 70, aucun indice n'a pu être calculé. Néanmoins, la sélection peut se faire uniquement sur base du nombre de paramètres, le set de variables minimisant ainsi le nombre de paramètres utilisés par le modèle final.

La validation croisée a démontré un modèle très hétérogène dans sa prédiction. Si l'indice AUC moyen n'est pas mauvais, son écart-type démontre une forte différence de qualité prédictive dans l'espace. En effet, les groupes 1 et 5 ont un AUC inférieur à 0.7 et donc mauvais, tandis que le groupe 3 en a un excellent. Le partitionnement en blocs ne permet pas de justifier cet écart. La contribution au modèle démontre une relative constance dans la contribution des milieux urbanisés. Néanmoins, la température minimale et le drainage participent dans des proportions différentes selon le groupe de validation, ce qui confirme le caractère hétérogène du modèle identifié par l'analyse de l'indice AUC.

Comparaison des cartes et futur de l'ambrosie en Wallonie

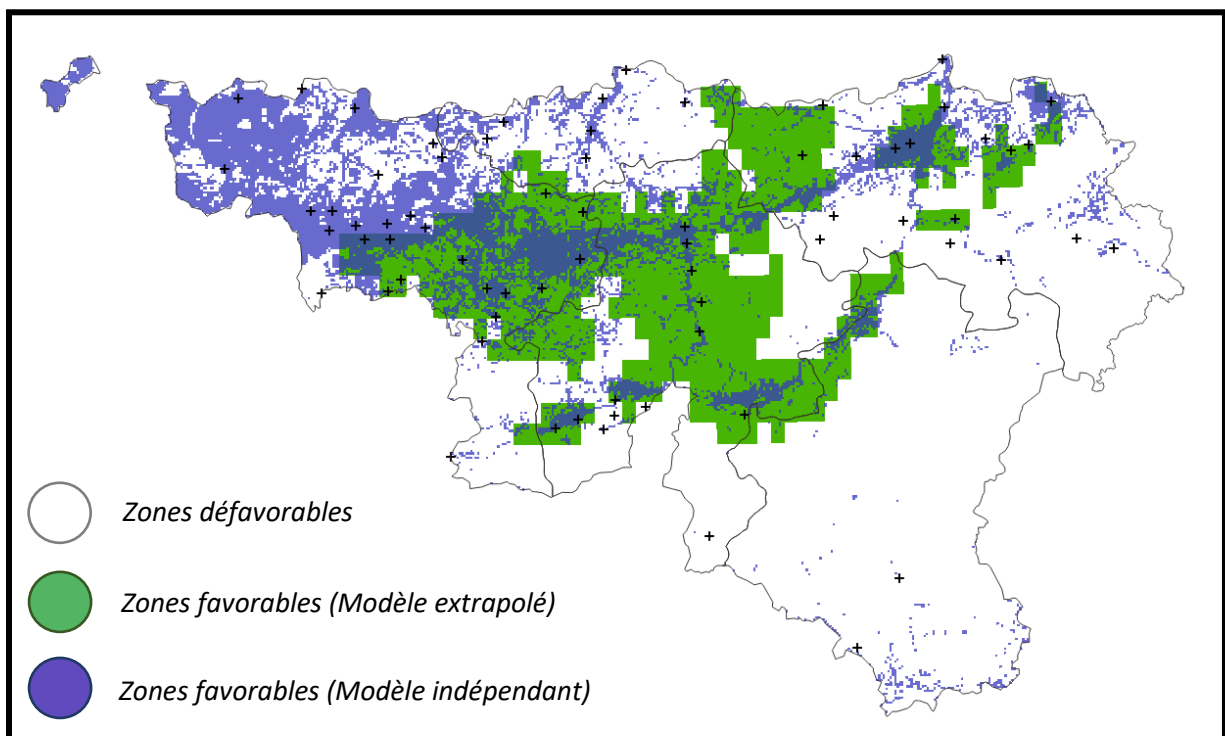


Figure 27 - Carte récapitulative des zones favorables et défavorables à l'ambrosie à feuilles d'armoise en région wallonne. Les zones favorables identifiées par le modèle extrapolé sont en vert tandis que les zones identifiées par le modèle indépendant sont en bleu. Les '+' représentent les points de présence utilisés pour la construction du modèle indépendant.

La superposition des deux cartes binaires permet de constater les similitudes entre les deux cartes (figure 27). En effet, à l'exception du nord-ouest de la région wallonne et de quelques zones dans les cantons de l'est et le sud de la province du Luxembourg, les zones identifiées sont similaires. Il est intéressant de constater cette similitude malgré l'indépendance des données en entrée, la différence des variables et la situation de l'ambrosie en Wallonie. Il existe cependant de nombreuses différences

entre les deux modèles. En effet, les variables liées au drainage et aux milieux artificiels ont fortement limité le modèle indépendant, élargissant la zone favorable au nord du Hainaut et au Brabant Wallon. De plus, la différence d'échelle de travail implique aussi une plus grande précision. Enfin, la faiblesse du jeu de données et l'historique d'invasion faussent le modèle indépendant, lorsque le modèle extrapolé jouit d'un jeu de données de présence plus robuste. Etant donné les dates de référence pour la construction des données rasters de Worldclim, il est néanmoins probable que le nord du Hainaut et le Brabant Wallon soient déjà dans l'enveloppe climatique optimale de l'ambrosie à feuilles d'armoise. De plus, le drainage des sols lui semble favorable dans ces régions, ce qui amènerait les provinces du Hainaut, de Namur et du Brabant Wallon en zone favorable. Il est difficile de prévoir quel est l'état actuel des données bioclimatiques, et donc l'enveloppe climatique actuelle de l'ambrosie. Néanmoins, selon les projections temporelles du modèle extrapolé et les résultats du modèle indépendant, il apparaît que la région wallonne est majoritairement en zone favorable à l'ambrosie à feuilles d'armoise, à l'exception de la province du Luxembourg et des cantons de l'Est.

Ces conclusions tendent à signifier une présence significative de l'ambrosie en région wallonne. En effet, présente depuis plusieurs décennies, la catégorisation d'une grande portion du territoire en zone favorable devrait être un facteur de choix pour la colonisation par l'espèce. Cependant, une vérification sur le terrain a démontré que bien que plusieurs populations aient été recensées, peu ont été repérées pendant la vérification de terrain de l'année 2020. Ce caractère fugace de l'ambrosie peut être expliqué par plusieurs phénomènes. D'une part, certaines populations se sont installées en zone non favorables, et n'ont certainement pas pu s'installer durablement dans ces régions. Il est aussi probable que les températures et les précipitations du début de l'année 2020 n'aient pas été assez propices à la sortie de dormance des graines d'ambrosie annuelle ou à son développement, mais il est peu probable qu'aucun plant n'émerge dans certaines populations recensées, quand d'autres endroits jouissent d'une importante densité, même s'ils sont rares. Enfin, d'autres phénomènes tels que des perturbations, la compétition interspécifique ou un potentiel effet Allee (une augmentation du taux de croissance avec la densité dans les petites populations) peuvent aussi expliquer cette relative absence de l'ambrosie dans le paysage actuel wallon.

L'identification de ces zones favorables mise en confrontation avec la vérification de terrain démontre que l'espèce est au début de sa colonisation. En effet, les données, bien que peu fiables, montrent que les présences sont majoritairement en zone urbanisée, pour des petites populations. Ces deux facteurs sont caractéristiques du début de l'invasion d'*Ambrosia artemisiifolia*. De plus, Bullock et al. (2012) caractérise les régions en début d'invasion comme petites et éphémères, ce qui est conforme aux populations identifiées en région wallonne.

Il est important de garder à l'esprit le caractère incertain des modèles sortants. En effet, ces modèles ne sont que des hypothèses à tester sur le terrain, une multitude de facteurs pouvant expliquer autrement la présence ou l'absence de l'espèce à un endroit donné. A plus petite échelle, des perturbations ou des différences de climat/environnement peuvent rendre des zones défavorables. Enfin, des barrières naturelles peuvent empêcher la propagation de l'espèce, bien que l'ambrosie soient peu concernée, particulièrement en Belgique. De plus, les multiples sources d'introduction amènent l'ambrosie sur tout le territoire wallon, avant de s'installer plus durablement en zone favorable (*réellement* favorable, et non spécialement en zone identifiée 'favorable' par le modèle). Ces sources d'introduction concernent majoritairement les graines pour oiseaux et la contamination des graines de tournesols, sources avérées d'introduction. Bien que ces sources soient avérées, aucune étude sur le sujet à l'échelle wallonne n'a encore vu le jour. Néanmoins, le rôle des champs de tournesols et des graines pour oiseaux dans la colonisation de l'espèce est avéré (Brandes and Nitzsche, 2006; B Fumanal et al., 2008; Pinke et al., 2013). Il serait cependant pertinent d'identifier les

autres sources d'introduction et dans quelles proportions elles sont importantes à l'échelle wallonne, afin d'améliorer le travail de prévention. En effet, si la présence de l'ambrosie en région wallonne ne semble pas être un danger actuellement, les modèles montrent clairement un environnement propice à son installation, qui le sera d'autant plus avec le réchauffement climatique. De par les inconvénients sanitaires et économiques qu'elle engendre, l'ambrosie à feuilles d'armoise doit donc faire l'objet d'un travail de prévention rigoureux afin d'éviter une installation durable et envahissante dans nos régions.

Ce travail de prévention passe avant tout par la sensibilisation auprès des acteurs liés à son introduction et sa dispersion, la constatation de l'état de l'ambrosie en région wallonne, dont la présente étude est une contribution et l'éradication des plants actuellement présents ainsi que leur suivi. Il existe plusieurs méthodes de destruction des plants, cependant le déracinement manuel pendant la phase végétative semble être le plus adapté à ce stade de colonisation de l'ambrosie. Cependant, une attention particulière devra être faite quant au signalement des populations, afin d'éviter la dispersion des populations actuelles.

Manipulations compromises par la situation sanitaire

Etant donné la crise sanitaire que le monde a subi lors de la période de la mise en œuvre de ce travail, il paraît pertinent de s'attarder sur les modifications apportées à l'étude initiale. En effet, bien que le sujet général soit similaire, le travail initial consistait en une modélisation plus expéditive, afin d'effectuer les autres objectifs correctement. Cette modélisation se serait alors faite avec Maxent, mais sans revue bibliographique plus poussée et sans les nombreux tests qui ont permis aux modélisations de ce travail de correspondre au mieux à l'espèce étudiée. Cependant, une extrapolation temporelle était prévue, le modèle extrapolé aurait donc été choisi.

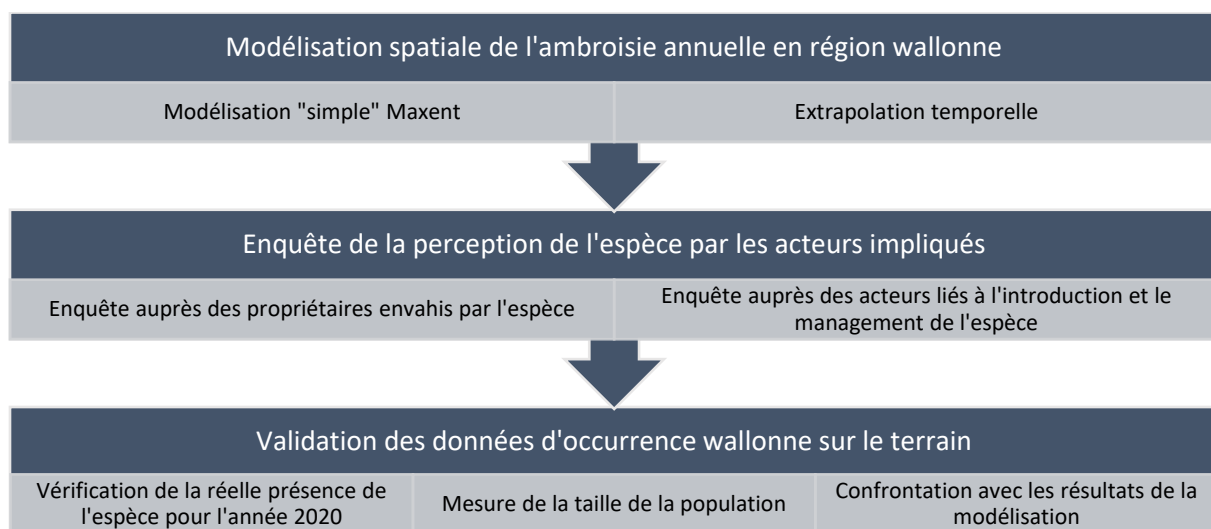


Figure 28 - Schéma des différents objectifs poursuivis par l'étude initialement

Outre la modélisation, les objectifs consistaient en une enquête de la perception de l'espèce et une validation des données sur le terrain (Figure 28). Les enquêtes sur la perception de l'espèce sont courantes et très utiles afin de cibler les points forts et points faibles perçus par les acteurs impliqués. Il est ainsi plus aisé de cibler la sensibilisation, et de mieux comprendre certains facteurs qui favorisent la colonisation de l'espèce exotique envahissante. Afin d'accentuer ce dernier point, une enquête spécifique était prévue pour les propriétaires des parcelles envahies. Une telle enquête aurait eu pour but de cibler les sources d'introduction favorisée en Wallonie, en comparaison des sources avérées d'autres pays envahis. Ces enquêtes auraient dû se faire lors des formations octroyées par l'Observatoire Wallon des Ambrosies, en complément de quelques interviews des propriétaires et

d'autres acteurs non présents en formation. La situation sanitaire a malheureusement fait reporter ces formations sous forme de webinaires, se prêtant moins aux enquêtes. De plus, ces webinaires eurent lieu bien plus tard, ne permettant pas d'optimiser le temps de travail.

Enfin, le dernier objectif consistait à faire valider les données de présence récoltées sur le terrain. De plus, par une mesure de l'aire de distribution et de la densité de chaque population, il aurait été possible d'obtenir la taille de chaque population et de générer une carte pondérée de la présence de l'ambrosie à feuilles d'armoise en région wallonne. De tels résultats auraient ainsi pu être confrontés à ceux de la modélisation. Les restrictions de déplacement et de travail imposés et la modification des objectifs de TFE ont contribué à écarter cet objectif, pour n'en faire qu'un sujet de discussion. En effet, cet objectif a bien été réalisé par l'Observatoire Wallon des Ambrosies, mais contraint par la présence d'une seule personne sur le terrain. Ainsi, seules deux journées de terrain ont été réalisées pour cette étude, afin de mieux comprendre les difficultés rencontrées lors de cette validation.

Les objectifs ont donc été modifiés pour ne concerner que la modélisation spatiale de l'ambrosie à feuilles d'armoise, avec un focus sur la littérature. D'une part, cela permettait une entière autonomie quant aux nouvelles mesures sanitaires, et d'autre part une compréhension poussée du sujet, permettant des modélisations plus qualitatives.

Conclusion

Depuis quelques dizaines d'années, les modélisations sont devenues des outils précieux dans les études de préservation de la biodiversité et de lutte contre les EEE. Grâce à ses faibles exigences en données d'entrée et l'accessibilité croissante de ces dernières, ces modèles sont devenus faciles d'accès et de plus en plus utilisés. Cependant, les différentes options disponibles afin de créer des modèles spécifiques à une espèce restent peu connues et sont rarement envisagées, malgré leur efficacité avérée. Dans le cadre des EEE, différentes méthodes ont été mises en évidence lors de cette étude, et ont mis en lumière les différents problèmes rencontrés lors de telles études.

Les deux travaux de modélisation sur la distribution de l'ambrosie en région wallonne ont montré l'affinité de l'ambrosie à feuilles d'armoise avec le territoire wallon. Seuls l'est de la province du Luxembourg et les cantons de l'Est semblent être en zone défavorable. Pour le cas de l'ambrosie à feuilles d'armoise, le modèle extrapolé a montré le rôle du climat dans sa répartition, et présente une certaine robustesse liée à la solidité du jeu de données d'entrée. Les variables contraignant le plus le modèle extrapolé sont les précipitations et la température des mois les plus chauds, ainsi que le coefficient de variation des précipitations. Le modèle indépendant, bien que souffrant d'un jeu de données de présence assez fragile, a présenté des similitudes avec le modèle extrapolé. Le modèle, contraint essentiellement par la distance aux chemins de fer suit le gradient de l'urbanisation sur la majorité du territoire. La proportion de drainage moyen a contraint les zones favorables dans le nord du Hainaut et l'ouest du Brabant Wallon. Enfin, les températures minimales du quartier le plus froid de l'année, conditionnant la sortie de dormance de l'espèce, est la dernière variable contribuant au modèle.

Malgré le faible recensement de l'ambrosie et les rares observations vérifiées, les modèles montrent que la région wallonne est un territoire propice à l'installation et la dispersion de l'ambrosie à feuilles d'armoise. De plus, le changement climatique devrait rapidement rendre la quasi-totalité de la région wallonne en zone favorable et donc favoriser son installation et sa propagation. Une vigilance accrue semble donc nécessaire dans les zones identifiées par les deux modèles, spécialement celles identifiées par le modèle extrapolé, qui a prouvé sa bonne transférabilité sur le territoire européen en identifiant les principaux foyers de l'ambrosie annuelle.

Les modèles construits dans cette étude peuvent encore être améliorés. L'intégration des variables de sols et des milieux anthropisés de la zone d'indigénat pourraient contribuer à un modèle de meilleure qualité prédictive. De plus, un échantillonnage encore plus représentatif des données de présence de la zone d'indigénat permettrait lui aussi de contribuer à un modèle plus qualitatif. Enfin, un recensement des points de présence plus rigoureux dans les zones européennes envahies pourrait permettre des modèles à échelle européenne exploitables.

De par les problèmes sanitaires et économiques qu'elle peut engendrer, l'ambrosie à feuilles d'armoise doit faire l'objet d'une attention particulière. Bien que l'espèce ne semble pas poser problème actuellement en région wallonne, les modèles ont montré que l'espèce pourrait se développer facilement dans le futur. Une prévention rigoureuse afin de limiter les sources d'introduction et favoriser la lutte est donc essentielle pour laisser cette espèce dans sa situation actuelle.

Bibliographie

- Aiello-Lammens, M.E., Boria, R.A., Radosavljevic, A., Vilela, B., Anderson, R.P., 2015. spThin: An R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography* 38, 541–545. <https://doi.org/10.1111/ecog.01132>
- Aikio, S., Duncan, R.P., Hulme, P.E., 2010. Herbarium records identify the role of long-distance spread in the spatial distribution of alien plants in New Zealand: Long-distance spread of alien plants. *Journal of Biogeography* 37, 1740–1751. <https://doi.org/10.1111/j.1365-2699.2010.02329.x>
- Albert, C.H., Yoccoz, N.G., Edwards, T.C., Graham, C.H., Zimmermann, N.E., Thuiller, W., 2010. Sampling in ecology and evolution - bridging the gap between theory and practice. *Ecography* 33, 1028–1037. <https://doi.org/10.1111/j.1600-0587.2010.06421.x>
- Alexander, J.M., Edwards, P.J., 2010. Limits to the niche and range margins of alien species. *Oikos* 119, 1377–1386. <https://doi.org/10.1111/j.1600-0706.2009.17977.x>
- Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS): Assessing the accuracy of distribution models. *Journal of Applied Ecology* 43, 1223–1232. <https://doi.org/10.1111/j.1365-2664.2006.01214.x>
- Anderson, R.P., Gonzalez, I., 2011. Species-specific tuning increases robustness to sampling bias in models of species distributions: An implementation with Maxent. *Ecological Modelling* 222, 2796–2811. <https://doi.org/10.1016/j.ecolmodel.2011.04.011>
- 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: Effect of study region on models of distributions. *Journal of Biogeography* 37, 1378–1393. <https://doi.org/10.1111/j.1365-2699.2010.02290.x>
- Andrade, A.F.A. de, Velazco, S.J.E., De Marco Júnior, P., 2020. ENMTML: An R package for a straightforward construction of complex ecological niche models. *Environmental Modelling & Software* 125, 104615. <https://doi.org/10.1016/j.envsoft.2019.104615>
- Araujo, M., New, M., 2007. Ensemble forecasting of species distributions. *Trends in Ecology & Evolution* 22, 42–47. <https://doi.org/10.1016/j.tree.2006.09.010>
- Araújo, M.B., Guisan, A., 2006. Five (or so) challenges for species distribution modelling. *J Biogeography* 33, 1677–1688. <https://doi.org/10.1111/j.1365-2699.2006.01584.x>
- Araujo, M.B., Pearson, R.G., Thuiller, W., Erhard, M., 2005. Validation of species-climate impact models under climate change. *Global Change Biol* 11, 1504–1513. <https://doi.org/10.1111/j.1365-2486.2005.01000.x>
- Araújo, M.B., Peterson, A.T., 2012. Uses and misuses of bioclimatic envelope modeling. *Ecology* 93, 1527–1539. <https://doi.org/10.1890/11-1930.1>
- Austin, M.P., 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling* 157, 101–118. [https://doi.org/10.1016/S0304-3800\(02\)00205-3](https://doi.org/10.1016/S0304-3800(02)00205-3)
- Barbet-Massin, M., Jiguet, F., Albert, C.H., Thuiller, W., 2012. Selecting pseudo-absences for species distribution models: how, where and how many?: How to use pseudo-absences in niche modelling? *Methods in Ecology and Evolution* 3, 327–338. <https://doi.org/10.1111/j.2041-210X.2011.00172.x>
- Barbet-Massin, M., Thuiller, W., Jiguet, F., 2010. How much do we overestimate future local extinction rates when restricting the range of occurrence data in climate suitability models? *Ecography* 33, 878–886. <https://doi.org/10.1111/j.1600-0587.2010.06181.x>
- Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S.P., Peterson, A.T., Soberón, J., Villalobos, F., 2011. The crucial role of the accessible area in ecological niche modeling and

- species distribution modeling. *Ecological Modelling* 222, 1810–1819.
<https://doi.org/10.1016/j.ecolmodel.2011.02.011>
- Baskin, J.M., Baskin, C.C., 1980. Ecophysiology of Secondary Dormancy in Seeds of *Ambrosia Artemisiifolia*. *Ecology* 61, 475–480. <https://doi.org/10.2307/1937410>
- Bassett, I.J., Crompton, C.W., 1975. THE BIOLOGY OF CANADIAN WEEDS.: 11. *Ambrosia artemisiifolia* L. and *A. psilostachya* DC. *Can. J. Plant Sci.* 55, 463–476. <https://doi.org/10.4141/cjps75-072>
- Bazzaz, F.A., 1979. The physiological ecology of plant succession. *Annu. Rev. Ecol. Syst.* 10, 351–371.
- Bazzaz, F.A., 1974. Ecophysiology of *Ambrosia Artemisiifolia*: A Successional Dominant. *Ecology* 55, 112–119. <https://doi.org/10.2307/1934623>
- Bedia, J., Herrera, S., Gutiérrez, J.M., 2013. Dangers of using global bioclimatic datasets for ecological niche modeling. Limitations for future climate projections. *Global and Planetary Change* 107, 1–12. <https://doi.org/10.1016/j.gloplacha.2013.04.005>
- Bocedi, G., Palmer, S.C.F., Pe'er, G., Heikkinen, R.K., Matsinos, Y.G., Watts, K., Travis, J.M.J., 2014. RangeShifter: a platform for modelling spatial eco-evolutionary dynamics and species' responses to environmental changes. *Methods in Ecology and Evolution* 5, 388–396. <https://doi.org/10.1111/2041-210X.12162>
- Boria, R.A., Olson, L.E., Goodman, S.M., Anderson, R.P., 2014. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecological Modelling* 275, 73–77. <https://doi.org/10.1016/j.ecolmodel.2013.12.012>
- Brandes, D., Nitzsche, J., 2006. Biology, introduction, dispersal, and distribution of common ragweed (*Ambrosia artemisiifolia* L.) with special regard to Germany 8.
- Brummer, T.J., Maxwell, B.D., Higgs, M.D., Rew, L.J., 2013. Implementing and interpreting local-scale invasive species distribution models. *Diversity Distrib.* 19, 919–932. <https://doi.org/10.1111/ddi.12043>
- Brunel, S., Branquart, E., Fried, G., Van Valkenburg, J., Brundu, G., Starfinger, U., Buholzer, S., Uludag, A., Joseffson, M., Baker, R., 2010. The EPPO prioritization process for invasive alien plants: EPPO prioritization for invasive alien plants. *EPPO Bulletin* 40, 407–422. <https://doi.org/10.1111/j.1365-2338.2010.02423.x>
- Buckley, L.B., Urban, M.C., Angilletta, M.J., Crozier, L.G., Rissler, L.J., Sears, M.W., 2010. Can mechanism inform species' distribution models?: Correlative and mechanistic range models. *Ecology Letters* no-no. <https://doi.org/10.1111/j.1461-0248.2010.01479.x>
- Bullock, J.M., Chapman, D., Schäfer, S., Roy, D., Girardello, M., Haynes, T., Beal, S., Wheeler, B., Dickie, I., Phang, Z., Tinch, R., Čivić, K., Delbaere, B., Jones-Walters, L., Hilbert, A., Schrauwen, A., Prank, M., Sofiev, M., Niemelä, S., Brough, C., 2012. Assessing and controlling the spread and the effects of common ragweed in Europe.
- Chapman, D., Pescott, O.L., Roy, H.E., Tanner, R., 2019. Improving species distribution models for invasive non-native species with biologically informed pseudo-absence selection. *J Biogeogr* 46, 1029–1040. <https://doi.org/10.1111/jbi.13555>
- Chapman, D.S., Haynes, T., Beal, S., Essl, F., Bullock, J.M., 2014. Phenology predicts the native and invasive range limits of common ragweed. *Glob Change Biol* 20, 192–202. <https://doi.org/10.1111/gcb.12380>
- Chapman, D.S., Makra, L., Albertini, R., Bonini, M., Páldy, A., Rodinkova, V., Šikoparija, B., Weryszko-Chmielewska, E., Bullock, J.M., 2016. Modelling the introduction and spread of non-native species: international trade and climate change drive ragweed invasion. *Glob Change Biol* 22, 3067–3079. <https://doi.org/10.1111/gcb.13220>
- Chauvel, B., Cadet, É., 2011. Introduction et dispersion d'une espèce envahissante: le cas de l'ambrosie à feuilles d'armoise (*Ambrosia artemisiifolia* L.) en France. *Acta Botanica Gallica* 158, 309–327. <https://doi.org/10.1080/12538078.2011.10516276>
- Chauvel, B., Dessaint, F., Cardinal-Legrand, C., Bretagnolle, F., 2006. The historical spread of *Ambrosia artemisiifolia* L. in France from herbarium records. *J Biogeography* 33, 665–673. <https://doi.org/10.1111/j.1365-2699.2005.01401.x>

- Chefaoui, R.M., Lobo, J.M., 2008. Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological Modelling* 210, 478–486. <https://doi.org/10.1016/j.ecolmodel.2007.08.010>
- Chen, H., Chen, L., Albright, T.P., 2007. Developing Habitat-suitability Maps of Invasive Ragweed (*Ambrosia artemisiifolia*L) in China Using GIS and Statistical Methods. *GIS for Health and the Environment, Lecture Notes in Geoinformation and Cartography* 105–121. https://doi.org/10.1007/978-3-540-71318-0_8
- Chollet, D. (Centre T.I. des O.M., Drieu, Y., Molines, J., Pauget, J., 1999. [Control of common ragweed [*Ambrosia artemisiifolia*]]. *Perspectives Agricoles (France)*.
- Colwell, R.K., Rangel, T.F., 2009. Hutchinson's duality: The once and future niche. *Proceedings of the National Academy of Sciences* 106, 19651–19658. <https://doi.org/10.1073/pnas.0901650106>
- Cunze, S., Leiblein, M.C., Tackenberg, O., 2013. Range Expansion of *Ambrosia artemisiifolia* in Europe Is Promoted by Climate Change. *ISRN Ecology* 2013, 1–9. <https://doi.org/10.1155/2013/610126>
- Dickerson, Tolbert, C., 1968. Studies on the germination, growth, development and control of common ragweed, *Ambrosia artemisiifolia* L.
- Dormann, C.F., Schymanski, S.J., Cabral, J., Chuine, I., Graham, C., Hartig, F., Kearney, M., Morin, X., Römermann, C., Schröder, B., Singer, A., 2012. Correlation and process in species distribution models: bridging a dichotomy: Bridging the correlation-process dichotomy. *J. Biogeogr.* 39, 2119–2131. <https://doi.org/10.1111/j.1365-2699.2011.02659.x>
- Drake, J.M., Bossenbroek, J.M., 2009. Profiling ecosystem vulnerability to invasion by zebra mussels with support vector machines. *Theor Ecol* 2, 189–198. <https://doi.org/10.1007/s12080-009-0050-8>
- Edwards, J.L., 2004. Research and Societal Benefits of the Global Biodiversity Information Facility. *BioScience* 54, 486. [https://doi.org/10.1641/0006-3568\(2004\)054\[0486:RASBOT\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2004)054[0486:RASBOT]2.0.CO;2)
- Elith, J., 2017. Predicting Distributions of Invasive Species, in: Robinson, A.P., Walshe, T., Burgman, M.A., Nunn, M. (Eds.), *Invasive Species*. Cambridge University Press, Cambridge, pp. 93–129. <https://doi.org/10.1017/9781139019606.006>
- Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., J. Phillips, S., Richardson, K., Scachetti-Pereira, R., E. Schapire, R., Soberón, J., Williams, S., S. Wisz, M., E. Zimmermann, N., 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29, 129–151. <https://doi.org/10.1111/j.2006.0906-7590.04596.x>
- Elith, J., Kearney, M., Phillips, S., 2010. The art of modelling range-shifting species: The art of modelling range-shifting species. *Methods in Ecology and Evolution* 1, 330–342. <https://doi.org/10.1111/j.2041-210X.2010.00036.x>
- Elith, J., Leathwick, J.R., 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annu. Rev. Ecol. Evol. Syst.* 40, 677–697. <https://doi.org/10.1146/annurev.ecolsys.110308.120159>
- Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., Yates, C.J., 2011. A statistical explanation of MaxEnt for ecologists: Statistical explanation of MaxEnt. *Diversity and Distributions* 17, 43–57. <https://doi.org/10.1111/j.1472-4642.2010.00725.x>
- Essl, F., Biró, K., Brandes, D., Broennimann, O., Bullock, J.M., Chapman, D.S., Chauvel, B., Dullinger, S., Fumanal, B., Guisan, A., Karrer, G., Kazinczi, G., Kueffer, C., Laitung, B., Lavoie, C., Leitner, M., Mang, T., Moser, D., Müller-Schärer, H., Petitpierre, B., Richter, R., Schaffner, U., Smith, M., Starfinger, U., Vautard, R., Vogl, G., von der Lippe, M., Follak, S., 2015. Biological Flora of the British Isles: *Ambrosia artemisiifolia*. *J Ecol* 103, 1069–1098. <https://doi.org/10.1111/1365-2745.12424>
- Essl, F., Dullinger, S., Kleinbauer, I., 2009. Changes in the spatio-temporal patterns and habitat preferences of *Ambrosia artemisiifolia* during its invasion of Austria. *Preslia* 81, 119–133.

- Fand, B.B., Shashank, P.R., Suroshe, S.S., Chandrashekar, K., Meshram, N.M., Timmanna, H.N., 2020. Invasion risk of the South American tomato pinworm *Tuta absoluta* (Meyrick) (Lepidoptera: Gelechiidae) in India: predictions based on MaxEnt ecological niche modelling. *Int J Trop Insect Sci.* <https://doi.org/10.1007/s42690-020-00103-0>
- Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37, 4302–4315. <https://doi.org/10.1002/joc.5086>
- Fourcade, Y., Engler, J.O., Rödder, D., Secondi, J., 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, e97122. <https://doi.org/10.1371/journal.pone.0097122>
- Franklin, J., 2010. Moving beyond static species distribution models in support of conservation biogeography: Moving beyond static species distribution models. *Diversity and Distributions* 16, 321–330. <https://doi.org/10.1111/j.1472-4642.2010.00641.x>
- Früh, L., Kampen, H., Kerkow, A., Schaub, G.A., Walther, D., Wieland, R., 2018. Modelling the potential distribution of an invasive mosquito species: comparative evaluation of four machine learning methods and their combinations. *Ecological Modelling* 388, 136–144. <https://doi.org/10.1016/j.ecolmodel.2018.08.011>
- Fumanal, B., 2007. Caractérisation des traits biologiques et des processus évolutifs d'une espèce envahissante en France: *Ambrosia artemisiifolia* L. Université de Bourgogne, France.
- Fumanal, B., Chauvel, B., Sabatier, A., Bretagnolle, F., 2007. Variability and Cryptic Heteromorphism of *Ambrosia artemisiifolia* Seeds: What Consequences for its Invasion in France? *Annals of Botany* 100, 305–313. <https://doi.org/10.1093/aob/mcm108>
- Fumanal, B., Gaudot, I., Bretagnolle, F., 2008. Seed-bank dynamics in the invasive plant, *Ambrosia artemisiifolia* L. *Seed Sci. Res.* 18, 101–114. <https://doi.org/10.1017/S0960258508974316>
- Fumanal, B., Girod, C., Fried, G., Bretagnolle, F., Chauvel, B., 2008. Can the large ecological amplitude of *Ambrosia artemisiifolia* explain its invasive success in France? *Weed Research* 48, 349–359. <https://doi.org/10.1111/j.1365-3180.2008.00627.x>
- Gallien, L., Douzet, R., Pratte, S., Zimmermann, N.E., Thuiller, W., 2012. Invasive species distribution models - how violating the equilibrium assumption can create new insights: Beyond the equilibrium assumption of SDMs. *Global Ecology and Biogeography* 21, 1126–1136. <https://doi.org/10.1111/j.1466-8238.2012.00768.x>
- Gallien, L., Münkemüller, T., Albert, C.H., Boulangeat, I., Thuiller, W., 2010. Predicting potential distributions of invasive species: where to go from here?: Predicting potential distributions of invasive species. *Diversity and Distributions* 16, 331–342. <https://doi.org/10.1111/j.1472-4642.2010.00652.x>
- GBIF.org, 2020. GBIF Occurrence Download. <https://doi.org/10.15468/dl.emus7z>
- Gentili, R., Gilardelli, F., Bona, E., Prosser, F., Selvaggi, A., Alessandrini, A., Martini, F., Nimis, P.L., Wilhalm, T., Adorni, M., Ardenghi, N.M.G., Barni, E., Bonafede, F., Bonini, M., Bouvet, D., Buffa, G., Ciappetta, S., Giordana, F., Faggi, G., Ghiani, A., Ghillani, L., Marcucci, R., Masin, R., Morelli, V., Montagnani, C., Montanari, S., Peccenini, S., Pellizzari, M., Romani, E., Saiani, D., Scortegagna, S., Sirotti, M., Truzzi, A., Vignodelli, M., Bagli, L., Fiandri, F., Siniscalco, C., Citterio, S., 2017. Distribution map of *Ambrosia artemisiifolia* L. (Asteraceae) in Italy. *Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology* 151, 381–386. <https://doi.org/10.1080/11263504.2016.1176966>
- Genton, B.J., Shykoff, J.A., Giraud, T., 2005. High genetic diversity in French invasive populations of common ragweed, *Ambrosia artemisiifolia*, as a result of multiple sources of introduction: GENETIC DIVERSITY IN INVASIVE COMMON RAGWEED. *Molecular Ecology* 14, 4275–4285. <https://doi.org/10.1111/j.1365-294X.2005.02750.x>
- Gormley, A.M., Forsyth, D.M., Griffioen, P., Lindeman, M., Ramsey, D.S.L., Scroggie, M.P., Woodford, L., 2011. Using presence-only and presence-absence data to estimate the current and potential distributions of established invasive species: Modelling invasive species

- distributions. *Journal of Applied Ecology* 48, 25–34. <https://doi.org/10.1111/j.1365-2664.2010.01911.x>
- Graham, C.H., Elith, J., Hijmans, R.J., Guisan, A., Townsend Peterson, A., Loisele, B.A., Anderson, R.P., Dudk, M., Ferrier, S., Huettmann, F., Leathwick, J., Lehmann, A., Li, J., Lohmann, L., Loisele, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J., Phillips, S., Richardson, K., Pereira, R.S., Schapire, R., Soberón, J., Williams, S., Wisz, M., Zimmermann, N., 2008. The influence of spatial errors in species occurrence data used in distribution models. *Journal of Applied Ecology* 45, 239–247. <https://doi.org/10.1111/j.1365-2664.2007.01408.x>
- Grinnell, J., 1917. The Niche-Relationships of the California Thrasher. *The Auk* 34, 427–433. <https://doi.org/10.2307/4072271>
- Gu, W., Swihart, R.K., 2004. Absent or undetected? Effects of non-detection of species occurrence on wildlife–habitat models. *Biological Conservation* 116, 195–203. [https://doi.org/10.1016/S0006-3207\(03\)00190-3](https://doi.org/10.1016/S0006-3207(03)00190-3)
- Guillemin, J.-P., Chauvel, B., 2011. Effects of the seed weight and burial depth on the seed behavior of common ragweed (*Ambrosia artemisiifolia*): Seed behavior of *Ambrosia artemisiifolia*. *Weed Biology and Management* 11, 217–223. <https://doi.org/10.1111/j.1445-6664.2011.00423.x>
- Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. *Ecol Letters* 8, 993–1009. <https://doi.org/10.1111/j.1461-0248.2005.00792.x>
- Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135, 147–186. [https://doi.org/10.1016/S0304-3800\(00\)00354-9](https://doi.org/10.1016/S0304-3800(00)00354-9)
- Guo, Q., Kelly, M., Graham, C.H., 2005. Support vector machines for predicting distribution of Sudden Oak Death in California. *Ecological Modelling* 182, 75–90. <https://doi.org/10.1016/j.ecolmodel.2004.07.012>
- Gusev, A.P., 2019. *Ambrosia artemisiifolia* in an Anthropogenic Landscape of Gomel Polesye. *Russ J Ecol* 50, 93–96. <https://doi.org/10.1134/S106741361901003X>
- Halvorsen, R., Mazzoni, S., Dirksen, J.W., Næsset, E., Gobakken, T., Ohlson, M., 2016. How important are choice of model selection method and spatial autocorrelation of presence data for distribution modelling by MaxEnt? *Ecological Modelling* 328, 108–118. <https://doi.org/10.1016/j.ecolmodel.2016.02.021>
- Harrison, S., 1997. HOW NATURAL HABITAT PATCHINESS AFFECTS THE DISTRIBUTION OF DIVERSITY IN CALIFORNIAN SERPENTINE CHAPARRAL 78, 9.
- Heckel, M.E., 1906. Sur l' *Ambrosia artemisiæfolia* L. et sa naturalisation en France. *Bulletin de la Société Botanique de France* 53, 600–620. <https://doi.org/10.1080/00378941.1906.10831212>
- Hernandez, P.A., Graham, C.H., Master, L.L., Albert, D.L., 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* 29, 773–785. <https://doi.org/10.1111/j.0906-7590.2006.04700.x>
- Hijmans, R.J., 2012. Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology* 93, 679–688. <https://doi.org/10.1890/11-0826.1>
- Hijmans, R.J., Graham, C.H., 2006. The ability of climate envelope models to predict the effect of climate change on species distributions. *Global Change Biol* 12, 2272–2281. <https://doi.org/10.1111/j.1365-2486.2006.01256.x>
- Hijmans, R.J., Phillips, S., Leathwick, J., Elith, J., 2017. dismo: Species distribution modeling. R package version 1, 1.1–4.
- Hilbert, D., 2001. The utility of artificial neural networks for modelling the distribution of vegetation in past, present and future climates. *Ecological Modelling* 146, 311–327. [https://doi.org/10.1016/S0304-3800\(01\)00323-4](https://doi.org/10.1016/S0304-3800(01)00323-4)
- Hutchinson, G.E., 1957. Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative Biology* 22, 415–427. <https://doi.org/10.1101/sqb.1957.022.01.039>

- Iturbide, M., Bedia, J., Herrera, S., del Hierro, O., Pinto, M., Gutiérrez, J.M., 2015. A framework for species distribution modelling with improved pseudo-absence generation. *Ecological Modelling* 312, 166–174. <https://doi.org/10.1016/j.ecolmodel.2015.05.018>
- Jarnevich, C.S., Stohlgren, T.J., Kumar, S., Morissette, J.T., Holcombe, T.R., 2015. Caveats for correlative species distribution modeling. *Ecological Informatics* 29, 6–15. <https://doi.org/10.1016/j.ecoinf.2015.06.007>
- Jeschke, J.M., Strayer, D.L., 2008. Usefulness of Bioclimatic Models for Studying Climate Change and Invasive Species. *Annals of the New York Academy of Sciences* 1134, 1–24. <https://doi.org/10.1196/annals.1439.002>
- Jiménez-Valverde, A., Lobo, J.M., Hortal, J., 2008. Not as good as they seem: the importance of concepts in species distribution modelling. *Diversity and Distributions* 14, 885–890. <https://doi.org/10.1111/j.1472-4642.2008.00496.x>
- Jiménez-Valverde, A., Peterson, A.T., Soberón, J., Overton, J.M., Aragón, P., Lobo, J.M., 2011. Use of niche models in invasive species risk assessments. *Biol Invasions* 13, 2785–2797. <https://doi.org/10.1007/s10530-011-9963-4>
- Kadmon, R., Farber, O., Danin, A., 2004. EFFECT OF ROADSIDE BIAS ON THE ACCURACY OF PREDICTIVE MAPS PRODUCED BY BIOCLIMATIC MODELS. *Ecological Applications* 14, 401–413. <https://doi.org/10.1890/02-5364>
- Kazinczi, G., Novák, R., Béres, I., 2008. Common ragweed (*Ambrosia artemisiifolia* L.) : a review with social regards to the results in Hungary. *Herbologia* 9, 55–91.
- Kearney, M., Phillips, B.L., Tracy, C.R., Christian, K.A., Betts, G., Porter, W.P., 2008. Modelling species distributions without using species distributions: the cane toad in Australia under current and future climates. *Ecography* 31, 423–434. <https://doi.org/10.1111/j.0906-7590.2008.05457.x>
- Kramer-Schadt, S., Niedballa, J., Pilgrim, J.D., Schröder, B., Lindenborn, J., Reinfelder, V., Stillfried, M., Heckmann, I., Scharf, A.K., Augeri, D.M., Cheyne, S.M., Hearn, A.J., Ross, J., Macdonald, D.W., Mathai, J., Eaton, J., Marshall, A.J., Semiadi, G., Rustam, R., Bernard, H., Alfred, R., Samejima, H., Duckworth, J.W., Breitenmoser-Wuersten, C., Belant, J.L., Hofer, H., Wilting, A., 2013. The importance of correcting for sampling bias in MaxEnt species distribution models. *Diversity Distrib.* 19, 1366–1379. <https://doi.org/10.1111/ddi.12096>
- Landolt, E., Bäumler, B., Ehrhardt, A., Hegg, O., Klötzli, F., Lämmler, W., Nobis, M., Rudmann-Maurer, K., Schweingruber, F., Theurillat, J.-P., Urmí, E., Vust, M., Wohlgemuth, T., 2010. Flora Indicativa – Ökologische Zeigerwerte und biologische Kennzeichen zur Flora der Schweiz und der Alpen.
- Lavoie, C., Jodoin, Y., De Merlis, A.G., 2007. How did common ragweed (*Ambrosia artemisiifolia* L.) spread in Québec? A historical analysis using herbarium records: Herbarium records and the spread of common ragweed. *Journal of Biogeography* 34, 1751–1761. <https://doi.org/10.1111/j.1365-2699.2007.01730.x>
- Liang, W., Papeş, M., Tran, L., Grant, J., Washington-Allen, R., Stewart, S., Wiggins, G., 2018. The effect of pseudo-absence selection method on transferability of species distribution models in the context of non-adaptive niche shift. *Ecological Modelling* 388, 1–9. <https://doi.org/10.1016/j.ecolmodel.2018.09.018>
- Lobo, J.M., Jiménez-Valverde, A., Hortal, J., 2010. The uncertain nature of absences and their importance in species distribution modelling. *Ecography* 33, 103–114. <https://doi.org/10.1111/j.1600-0587.2009.06039.x>
- Lobo, J.M., Jiménez-Valverde, A., Real, R., 2008. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecol Biogeography* 17, 145–151. <https://doi.org/10.1111/j.1466-8238.2007.00358.x>
- Loiselle, B.A., Jørgensen, P.M., Consiglio, T., Jiménez, I., Blake, J.G., Lohmann, L.G., Montiel, O.M., 2007. Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? *Journal of Biogeography* 0, 070908043732003–???. <https://doi.org/10.1111/j.1365-2699.2007.01779.x>

- Lustig, A., Worner, S.P., Pitt, J.P.W., Doscher, C., Stouffer, D.B., Senay, S.D., 2017. A modeling framework for the establishment and spread of invasive species in heterogeneous environments. *Ecol Evol* 7, 8338–8348. <https://doi.org/10.1002/ece3.2915>
- Lütolf, M., Kienast, F., Guisan, A., 2006. The ghost of past species occurrence: improving species distribution models for presence-only data: Improving species distribution models. *Journal of Applied Ecology* 43, 802–815. <https://doi.org/10.1111/j.1365-2664.2006.01191.x>
- Mau-Crimmins, T.M., Schussman, H.R., Geiger, E.L., 2006. Can the invaded range of a species be predicted sufficiently using only native-range data? *Ecological Modelling* 193, 736–746. <https://doi.org/10.1016/j.ecolmodel.2005.09.002>
- McPHERSON, J.M., Jetz, W., Rogers, D.J., 2004. The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact?: Species' range and distribution model accuracy. *Journal of Applied Ecology* 41, 811–823. <https://doi.org/10.1111/j.0021-8901.2004.00943.x>
- Milakovic, I., Fiedler, K., Karrer, G., 2014. Fine-tuning of a mowing regime, a method for the management of the invasive plant, *Ambrosia artemisiifolia*, at different population densities: Management of common ragweed. *Weed Biology and Management* 14, 232–241. <https://doi.org/10.1111/wbm.12051>
- Milakovic, I., Karrer, G., 2016. The influence of mowing regime on the soil seed bank of the invasive plant *Ambrosia artemisiifolia* L. *NB* 28, 39–49. <https://doi.org/10.3897/neobiota.28.6838>
- Montagnani, C., Gentili, R., Smith, M., Guarino, M.F., Citterio, S., 2017. The Worldwide Spread, Success, and Impact of Ragweed (*Ambrosia* spp.). *Critical Reviews in Plant Sciences* 36, 139–178. <https://doi.org/10.1080/07352689.2017.1360112>
- Monty, A., Mahy, G., 2009. Evolution des traits d'histoire de vie lors des invasions végétales. *Biotechnol. Agron. Soc. Environ.* 10.
- Morin, X., Thuiller, W., 2009. Comparing niche- and process-based models to reduce prediction uncertainty in species range shifts under climate change. *Ecology* 90, 1301–1313. <https://doi.org/10.1890/08-0134.1>
- Muscarella, R., Galante, P.J., Soley-Guardia, M., Boria, R.A., Kass, J.M., Uriarte, M., Anderson, R.P., 2014. ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods in Ecology and Evolution* 5, 1198–1205. <https://doi.org/10.1111/2041-210x.12261>
- Nitzsche, J., 2010. *Ambrosia artemisiifolia* L. (Beifuß-Ambrosie) in Deutschland: Biologie der Art, Konkurrenzverhalten und Monitoring (PhD Thesis).
- Nix, H., Busby, J., 1986. BIOCLIM, a bioclimatic analysis and prediction system. Division of Water and Land Resources: Canberra.
- Ortmans, W., Mahy, G., Chauvel, B., Monty, A., 2016a. Performance variation of common ragweed (*Ambrosia artemisiifolia* L.) across invasion levels in Western Europe. *Flora - Morphology, Distribution, Functional Ecology of Plants* 220, 134–141. <https://doi.org/10.1016/j.flora.2016.03.003>
- Ortmans, W., Mahy, G., Monty, A., 2017. Northern range edge equilibrium of *Ambrosia artemisiifolia* L. not achieved in Western Europe. *Biotechnol. Agron. Soc. Environ.* 11.
- Ortmans, W., Mahy, G., Monty, A., 2016b. Effects of seed traits variation on seedling performance of the invasive weed, *Ambrosia artemisiifolia* L. *Acta Oecologica* 71, 39–46. <https://doi.org/10.1016/j.actao.2016.01.008>
- Ozaslan, C., Onen, H., Farooq, S., Gunal, H., Akyol, N., 2016. Common ragweed: An emerging threat for sunflower production and human health in Turkey: Ragweed invasion in Turkey. *Weed Biology and Management* 16, 42–55. <https://doi.org/10.1111/wbm.12093>
- Peterson, A.T., 2006a. Uses and Requirements of Ecological Niche Models and Related Distributional Models. *Biodiv. Inf.* 3. <https://doi.org/10.17161/bi.v3i0.29>
- Peterson, A.T., 2006b. Uses and Requirements of Ecological Niche Models and Related Distributional Models. *Biodiv. Inf.* 3. <https://doi.org/10.17161/bi.v3i0.29>

- Peterson, A.T., Soberón, J., 2012. Species Distribution Modeling and Ecological Niche Modeling: Getting the Concepts Right. *NatCon* 10, 102–107. <https://doi.org/10.4322/natcon.2012.019>
- Petitpierre, B., Kueffer, C., Broennimann, O., Randin, C., Daehler, C., Guisan, A., 2012. Climatic Niche Shifts Are Rare Among Terrestrial Plant Invaders. *Science* 335, 1344–1348. <https://doi.org/10.1126/science.1215933>
- Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190, 231–259. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>
- Phillips, S.J., Dudík, M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31, 161–175. <https://doi.org/10.1111/j.0906-7590.2008.5203.x>
- Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., Ferrier, S., 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications* 19, 181–197. <https://doi.org/10.1890/07-2153.1>
- Pierce, D.W., Barnett, T.P., Santer, B.D., Gleckler, P.J., 2009. Selecting global climate models for regional climate change studies. *Proceedings of the National Academy of Sciences* 106, 8441–8446. <https://doi.org/10.1073/pnas.0900094106>
- Pinke, G., Karácsony, P., Botta-Dukát, Z., Czúcz, B., 2013. Relating *Ambrosia artemisiifolia* and other weeds to the management of Hungarian sunflower crops. *J Pest Sci* 86, 621–631. <https://doi.org/10.1007/s10340-013-0484-z>
- Pinke, G., Karácsony, P., Czúcz, B., Botta-Dukát, Z., 2011. Environmental and land-use variables determining the abundance of *Ambrosia artemisiifolia* in arable fields in Hungary. *Preslia* 83, 219–235.
- Pinke, G., Kolejanisz, T., Vér, A., Nagy, K., Milics, G., Schlögl, G., Bede-Fazekas, Á., Botta-Dukát, Z., Czúcz, B., 2019. Drivers of *Ambrosia artemisiifolia* abundance in arable fields along the Austrian-Hungarian border. *Preslia* 91, 369–389. <https://doi.org/10.23855/preslia.2019.369>
- Pulliam, H.R., 2000. On the relationship between niche and distribution. *Ecol Letters* 3, 349–361. <https://doi.org/10.1046/j.1461-0248.2000.00143.x>
- Pyšek, P., Richardson, D.M., Rejmánek, M., Webster, G.L., Williamson, M., Kirschner, J., 2004. Alien plants in checklists and floras: Towards better communication between taxonomists and ecologists. *Taxon* 53, 131–143. <https://doi.org/10.2307/4135498>
- QGIS.org, 2020. QGIS Geographic Information System. Open Source Geospatial Foundation Project.
- Qin, Z., DiTommaso, A., Wu, R.S., Huang, H.Y., 2014. Potential distribution of two *Ambrosia* species in China under projected climate change. *Weed Res* 54, 520–531. <https://doi.org/10.1111/wre.12100>
- R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.
- Radosavljevic, A., Anderson, R.P., 2014. Making better Maxent models of species distributions: complexity, overfitting and evaluation. *J. Biogeogr.* 41, 629–643. <https://doi.org/10.1111/jbi.12227>
- Reznik, S.Y., 2009. Common ragweed (*Ambrosia artemisiifolia* L.) in Russia: spread, distribution, abundance, harmfulness and control measures 10.
- Ricciardi, A., Blackburn, T.M., Carlton, J.T., Dick, J.T.A., Hulme, P.E., Iacarella, J.C., Jeschke, J.M., Liebhold, A.M., Lockwood, J.L., MacIsaac, H.J., Pyšek, P., Richardson, D.M., Ruiz, G.M., Simberloff, D., Sutherland, W.J., Wardle, D.A., Aldridge, D.C., 2017. Invasion Science: A Horizon Scan of Emerging Challenges and Opportunities. *Trends in Ecology & Evolution* 32, 464–474. <https://doi.org/10.1016/j.tree.2017.03.007>
- Richardson, D.M., Pyšek, P., Rejmanek, M., Barbour, M.G., Panetta, F.D., West, C.J., 2000. Naturalization and invasion of alien plants: concepts and definitions. *Diversity & Distributions* 6, 93–107. <https://doi.org/10.1046/j.1472-4642.2000.00083.x>

- Rödger, D., Schmidlein, S., Veith, M., Lötters, S., 2009. Alien Invasive Slider Turtle in Unpredicted Habitat: A Matter of Niche Shift or of Predictors Studied? PLoS ONE 4, e7843. <https://doi.org/10.1371/journal.pone.0007843>
- Rondinini, C., Wilson, K.A., Boitani, L., Grantham, H., Possingham, H.P., 2006. Tradeoffs of different types of species occurrence data for use in systematic conservation planning: Species data for conservation planning. Ecology Letters 9, 1136–1145. <https://doi.org/10.1111/j.1461-0248.2006.00970.x>
- Schröder, B., 2008. Challenges of species distribution modeling belowground. J. Plant Nutr. Soil Sci. 171, 325–337. <https://doi.org/10.1002/jpln.200700027>
- Schurr, F.M., Pagel, J., Cabral, J.S., Groeneveld, J., Bykova, O., O’Hara, R.B., Hartig, F., Kissling, W.D., Linder, H.P., Midgley, G.F., Schröder, B., Singer, A., Zimmermann, N.E., 2012. How to understand species’ niches and range dynamics: a demographic research agenda for biogeography: A demographic research agenda for biogeography. J. Biogeogr. 39, 2146–2162. <https://doi.org/10.1111/j.1365-2699.2012.02737.x>
- Sexton, J.P., McIntyre, P.J., Angert, A.L., Rice, K.J., 2009. Evolution and Ecology of Species Range Limits. Annu. Rev. Ecol. Evol. Syst. 40, 415–436. <https://doi.org/10.1146/annurev.ecolsys.110308.120317>
- Shabani, F., Kumar, L., Ahmadi, M., 2016. A comparison of absolute performance of different correlative and mechanistic species distribution models in an independent area. Ecol Evol 6, 5973–5986. <https://doi.org/10.1002/ece3.2332>
- Shrestha, A., Roman, E., Thomas, A., Swanton, C., 1999. Modeling germination and shoot-radicle elongation of *Ambrosia artemisiifolia*. Weed Science 47, 557–562. <https://doi.org/10.1017/S0043174500092262>
- Skálová, H., Moravcová, L., Dixon, A.F.G., Kindlmann, P., Pyšek, P., 2015. Effect of temperature and nutrients on the growth and development of seedlings of an invasive plant. AoB PLANTS 7. <https://doi.org/10.1093/aobpla/plv044>
- Smith, M., Cecchi, L., Skjøth, C.A., Karrer, G., Šikoparija, B., 2013. Common ragweed: A threat to environmental health in Europe. Environment International 61, 115–126. <https://doi.org/10.1016/j.envint.2013.08.005>
- Soberón, J.M., 2010. Niche and area of distribution modeling: a population ecology perspective. Ecography 33, 159–167. <https://doi.org/10.1111/j.1600-0587.2009.06074.x>
- Steiner, F.M., Schlick-Steiner, B.C., VanDerWal, J., Reuther, K.D., Christian, E., Stauffer, C., Suarez, A.V., Williams, S.E., Crozier, R.H., 2008. Combined modelling of distribution and niche in invasion biology: a case study of two invasive *Tetramorium* ant species: Distribution and niche in invasion modelling. Diversity and Distributions 14, 538–545. <https://doi.org/10.1111/j.1472-4642.2008.00472.x>
- Sullivan, M.J.P., Davies, R.G., Reino, L., Franco, A.M.A., 2012. Using dispersal information to model the species-environment relationship of spreading non-native species: *Modelling dispersal-limited species*. Methods in Ecology and Evolution 3, 870–879. <https://doi.org/10.1111/j.2041-210X.2012.00219.x>
- Thuiller, W., Lafourcade, B., Engler, R., Araújo, M.B., 2009. BIOMOD - a platform for ensemble forecasting of species distributions. Ecography 32, 369–373. <https://doi.org/10.1111/j.1600-0587.2008.05742.x>
- Václavík, T., Kupfer, J.A., Meentemeyer, R.K., 2012. Accounting for multi-scale spatial autocorrelation improves performance of invasive species distribution modelling (iSDM): Multi-scale spatial autocorrelation and invasive species distribution models. Journal of Biogeography 39, 42–55. <https://doi.org/10.1111/j.1365-2699.2011.02589.x>
- Václavík, T., Meentemeyer, R.K., 2009. Invasive species distribution modeling (iSDM): Are absence data and dispersal constraints needed to predict actual distributions? Ecological Modelling 220, 3248–3258. <https://doi.org/10.1016/j.ecolmodel.2009.08.013>
- Valavi, R., Elith, J., Lahoz-Monfort, J.J., Guillera-Aroita, G., 2019. blockCV : An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species

- distribution models. *Methods in Ecology and Evolution* 10, 225–232.
<https://doi.org/10.1111/2041-210X.13107>
- VanDerWal, J., Shoo, L.P., Graham, C., Williams, S.E., 2009. Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know? *Ecological Modelling* 220, 589–594. <https://doi.org/10.1016/j.ecolmodel.2008.11.010>
- Warren, D.L., Seifert, S.N., 2011. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications* 21, 335–342. <https://doi.org/10.1890/10-1171.1>
- Warren, D.L., Wright, A.N., Seifert, S.N., Shaffer, H.B., 2014. Incorporating model complexity and spatial sampling bias into ecological niche models of climate change risks faced by 90 California vertebrate species of concern. *Diversity Distrib.* 20, 334–343.
<https://doi.org/10.1111/ddi.12160>
- West, A.M., Kumar, S., Brown, C.S., Stohlgren, T.J., Bromberg, J., 2016. Field validation of an invasive species Maxent model. *Ecological Informatics* 36, 126–134.
<https://doi.org/10.1016/j.ecoinf.2016.11.001>
- Willemsen, R., 1975. Effect of Stratification Temperature and Germination Temperature on Germination and the Induction of Secondary Dormancy in Common Ragweed Seeds. *American Journal of Botany* 62, 1. <https://doi.org/10.2307/2442073>
- Williamson, M., 1996. *Biological Invasions*. Chapman and Hall, London.
- Yan, H., Feng, Lei, Zhao, Y., Feng, Li, Zhu, C., Qu, Y., Wang, H., 2020. Predicting the potential distribution of an invasive species, *Erigeron canadensis* L., in China with a maximum entropy model. *Global Ecology and Conservation* 21, e00822.
<https://doi.org/10.1016/j.gecco.2019.e00822>
- Zaniewski, A.E., Lehmann, A., Overton, J.M., 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modelling* 157, 261–280. [https://doi.org/10.1016/S0304-3800\(02\)00199-0](https://doi.org/10.1016/S0304-3800(02)00199-0)
- Ziska, L.H., Caulfield, F.A., 2000. Rising CO₂ and pollen production of common ragweed (*Ambrosia artemisiifolia* L.), a known allergy-inducing species: implications for public health. *Functional Plant Biol.* 27, 893. <https://doi.org/10.1071/PP00032>

Annexe

Annexe 1 : Tableau comparatif de plusieurs algorithmes permettant de créer un modèle de distribution d'espèce. Tableau tiré de Elith (2017).

General class	Model (abbreviation)	Species data	Partial plots for effect on response	Comment
Expert model	Habitat suitability index (HSI)	Expert	Yes	Use expert knowledge for shape of species response
Expert model	Expert	Expert/presence	No	Use expert knowledge to select variables and perhaps to inform about presence
Climate envelope	BIOCLIM	Presence	No	Delimits climate envelope only using presence data, sometimes using percentiles; prediction from most extreme (limiting) variable
Machine learning	One-class support vector machines	Presence	No	Few uses, but being included in some ensembles
Factor analysis	Ecological niche factor analysis (ENFA)	Presence-background	No	Also known as Biomapper
Machine learning	Genetic algorithm for ruleset production (GARP)	Presence-background	No	Widely used; final model is an average over best selected rules
Machine Learning	Maximum entropy (MaxEnt)	Presence-background	Yes	Widely used; complexity of model can be adjusted by choice of features and adjusting regularisation
Regression	Generalised linear models (GLMs) or generalised additive models (GAMs)	Various	Yes	Statistical regression methods; generalised additive models allow smoothed data-driven functions
Regression	Non-parametric multiplicative regression	Various	Yes	Implemented in Hyperniche; only found invasive examples use invaded range data
Machine Learning	Decision tree	Various	Yes	Also known as classification and regression trees; more often used for decision analysis (e.g. on whether species will become invasive or not)
Machine Learning	Ensembles of trees: boosted regression trees (BRT), or random forests (RF)	Various	Some	Most invasive species examples are within ensembles; automatically model interactions unless stumps used
Machine Learning	Artificial neural nets	Various	Some	One of the earliest machine learning methods to be used in species modelling; regarded as a good general purpose algorithm
Ensembles	Ensembles of any type of models	Not applicable	No	Several examples emerging, with varied approaches for selecting the component models