

## GUIDE FRANCOPHONE POUR LA MODELISATION DE NICHES ECOLOGIQUES

ANAÏS VIGNOLES<sup>1\*</sup>

<sup>1</sup>UMR-5199 PACEA, Université de Bordeaux, Bâtiment B2, Allée Geoffroy Saint-Hilaire, CS 50023, 33615 Pessac Cedex (France)

*Résumé.* La modélisation corrélacionnelle de niches écologiques (en anglais « *Ecological Niche Modeling* »; ENM) est un ensemble de méthodes populaire dans le champ de l'écologie de la distribution d'espèces et est employée pour une multitude d'applications. Si le cadre conceptuel et méthodologique de l'ENM a été largement décrit dans la littérature, il n'existe pas de synthèse exhaustive en langue française. Dans cet article, j'expose les bases théoriques de l'ENM à travers un historique du concept de niche écologique et ses implications pour l'étude de la distribution macro-géographique des espèces. Je décris ensuite les différentes étapes d'une étude ENM, en insistant tout d'abord sur l'importance de contrôler la qualité des données d'entrées. Différentes préconisations concernant le choix des algorithmes, la calibration et l'évaluation des modèles ainsi que les analyses postérieures, telles que les comparaisons de niches ou le transfert à d'autres périodes/régions, sont présentées. J'insiste en particulier sur 1/ le fonctionnement de l'algorithme Maxent – l'algorithme le plus usité dans la littérature actuellement – et la nécessité d'un processus de réglage de ses paramètres, 2/ l'importance du choix de l'aire de calibration **M**, 3/ la nécessité de prendre en compte les environnements accessibles (associés à l'aire de calibration **M**) dans le transfert et la comparaison des modèles, et 4/ l'importance d'évaluer et de présenter la variabilité des résultats en fonction de choix méthodologiques à différentes étapes (partitionnement des données d'occurrences, choix d'un modèle climatique, choix de l'algorithme, choix de l'aire de calibration, etc.). En conclusion, je rappelle l'importance d'ancrer toute étude employant l'ENM dans un cadre théorique et méthodologique clair et explicite afin de garantir la pertinence des interprétations ultérieures.

*Mots-clés:* modélisation de niches écologiques; bonnes pratiques; cadre conceptuel; calibration et évaluation des modèles; transfert de modèles; comparaison de modèles

*Abstract:* Correlational ecological niche modeling (ENM) is a popular group of methods in the field of distributional ecology and is employed for a variety of applications. Although the conceptual and methodological framework of ENM has been widely described in the literature, there is still no exhaustive synthesis of it in the French language. In this article, theoretical bases of ENM are exposed through a history of the concept of ecological niche as well as its implications for the study of species macroscale distributions. Then, the different steps of ENM are described, emphasizing on the importance of controlling the quality of input data. Various recommendations concerning algorithm choice, model calibration and evaluation as well post-modeling analyses, such as niche comparison and transfer to other periods/regions, are presented. Particular emphasis is placed on 1/ the operation of Maxent – the most used algorithm in the literature today – and the need for parameter tuning prior modeling, 2/ the importance of the choice of the **M** calibration area, 3/ the need to take into account accessible environments (associated with the **M** calibration area) for model transfer and comparison, and 4/ the importance of evaluating and presenting the variability of models resulting from methodological choices at different stages (occurrence data partitioning, choice of a climate model, choice of algorithm, choice of the calibration area, etc.). To conclude, contextualizing any ENM study in a clear and explicit theoretical and methodological framework is paramount to ensure the pertinence of subsequent interpretations.

*Key words:* ecological niche modeling; good practices; conceptual framework; model calibration and evaluation; model transfer; model comparison

---

\*Corresponding author: [anaïs.l.vignoles@gmail.com](mailto:anaïs.l.vignoles@gmail.com)

La modélisation de niches écologiques (ENM) désigne un ensemble de méthodes corrélatives visant à reconstituer la niche écologique d'un taxon ou d'une population. Les modèles obtenus s'appuient sur des corrélations entre des occurrences connues pour ledit taxon et des variables environnementales présentées sous forme de cartes à plus ou moins haute résolution. Ils sont utilisés pour une grande diversité d'applications et de problématiques, telles que la découverte de nouvelles espèces ou populations (e.g., Raxworthy *et al.*, 2003; de Siqueira *et al.*, 2009; Peterson et Navarro-Sigüenza, 2009), la planification des politiques de conservation de la biodiversité (e.g., Sohn *et al.*, 2013; Sobral-Souza *et al.*, 2021), les dynamiques spatiales d'espèces invasives (e.g., Zhu *et al.*, 2012; Escobar *et al.*, 2014; Alkische *et al.*, 2020; Nuñez-Penichet *et al.*, 2021a), les mécanismes de macroévolution des espèces (e.g., Saube *et al.*, 2019), la paléocologie et paléo-distribution d'espèces fossiles (e.g., Myers *et al.*, 2015; Gibert, Vignoles *et al.*, 2022), la diffusion de maladies *via* des vecteurs animaux (e.g., Sweeney *et al.*, 2006; Escobar *et al.*, 2017; Marques *et al.*, 2020, 2021), l'impact du changement climatique sur les espèces animales ou végétales (e.g., Warren *et al.*, 2014; Ashraf *et al.*, 2017), ou encore les relations culture-environnement chez les chasseurs-cueilleurs du Paléolithique (e.g., Banks *et al.*, 2006, 2009, 2011, 2021; Vignoles *et al.*, 2021). Dans les années 2000, le nombre d'études employant l'ENM a explosé (Lobo *et al.*, 2010). La démocratisation progressive de cette méthodologie a malheureusement donné suite à de nombreuses applications erronées d'un point de vue conceptuel ou insuffisamment robustes d'un point de vue méthodologique. Si aujourd'hui de nombreux articles publiés en langue anglaise proposent des clarifications terminologiques (e.g., Soberón et Peterson, 2005; Soberón et Nakamura, 2009; Peterson et Soberón, 2012; Warren, 2012; Araújo et Peterson, 2012) et des guides résumant les principales étapes et les bonnes pratiques associées en ENM (e.g., Peterson *et al.*, 2011; Feng *et al.*, 2019a; Araújo *et al.*, 2019; Sillero et Barbosa, 2020; Sillero *et al.*, 2021), il n'existe aucune référence récapitulant les concepts et les précautions nécessaires à l'emploi de cette approche en langue française (voir cependant Pierrat, 2011; Antunes, 2015; Vignoles, 2021).

L'application de l'ENM se divise en quatre principales étapes. La première (1) est de se positionner clairement vis-à-vis du cadre théorique de cette approche. Les concepts et la terminologie

employés posent en effet le cadre de l'étude; ils délimitent le type de problématique qui peut être résolu par l'ENM, et donnent aux utilisateurs une référence pour le protocole de modélisation et pour correctement interpréter les modèles. La seconde étape (2) vise à recueillir les données à l'origine des modèles de niches. Celles-ci sont de deux natures: tout d'abord, les données d'occurrences permettent de décrire la répartition géographique de l'espèce ou de la population considérée, par l'identification de localités dans lesquelles des spécimens ont été observés à un instant *t*. Ensuite, les données environnementales représentent l'environnement dans lequel les spécimens ont évolué, par le biais de cartes de répartition de différentes variables (e.g., climat, topographie, etc.). Ces données doivent être sélectionnées avec soin, pour éviter le phénomène de « *Garbage-in, garbage-out* »: l'emploi de données trop éloignées de la réalité à ce stade se répercutera nécessairement sur les modèles de niches, qui seront alors moins pertinents pour répondre aux problématiques de départ. La troisième étape (3) est la création de modèles par la mise en relation de ces deux types de données *via* des méthodes statistiques ou la modélisation de surfaces de réponse. Les algorithmes pouvant être utilisés sont nombreux et la pertinence de leur utilisation dépend du type de données employées (e.g., données d'occurrences de type présence-seule, présence-absence, etc.) et des objectifs de l'étude (e.g., modéliser une distribution potentielle, visualiser des volumes dans un espace environnemental, etc.). Enfin, la dernière étape (4) consiste à projeter le modèle dans un nouveau contexte environnemental – comme une autre région ou une autre période. À ce stade, des modèles peuvent être comparés dans l'objectif d'évaluer leur degré de similarité ou de recouvrement. Cette dernière étape est particulièrement intéressante pour prédire l'impact de changements environnementaux sur une population (e.g., Ashraf *et al.*, 2017) ou pour estimer la capacité dispersive d'une espèce potentiellement invasive dans une région donnée (e.g., Nuñez-Penichet *et al.*, 2021a).

Dans cet article, je propose une synthèse des concepts et de la terminologie adéquate à l'ENM, ainsi que les principales précautions à prendre en compte lors de la mise en œuvre de cette approche à un contexte d'application, quel qu'il soit – depuis le choix des données jusqu'à la création, le transfert et la comparaison des modèles.

## CONCEPTS ET TERMINOLOGIE

*Niche Grinnellienne et niche Eltonienne*

La première définition du concept de niche est communément attribuée au zoologue américain J. Grinnell, dans un article portant sur « *Les relations de niche du moqueur de Californie* » (Grinnell, 1917; Chase et Leibold, 2003). Dans cette publication, l'auteur met en relation la distribution géographique particulière du moqueur de Californie (*Toxostoma redivivum*), avec sa « niche », qu'il définit comme l'expression géographique de ses prérequis environnementaux (en anglais, « *requirements* »), telles que ses tolérances physiologiques ou ses habitudes alimentaires. Dans les années 1930, une vision de la « niche » assez différente est proposée par l'écologue anglais C. Elton, considéré comme l'un des pionniers de l'écologie des populations et communautés. Il utilise ce terme pour désigner le rôle fonctionnel d'un animal au sein de la chaîne trophique (Elton, 1927). Cette définition se concentre alors sur l'impact d'un organisme sur son environnement, par exemple par le biais de la consommation de ressources ou de la compétition avec d'autres organismes (« *food and enemies* »). Cette double signification du mot « niche » – « *niche as requirements* versus *niche as impacts* » – est probablement à l'origine d'une confusion autour de ce concept, qui provoque d'importantes controverses épistémologiques dans la seconde moitié du XX<sup>ème</sup> siècle, allant jusqu'à son quasi-abandon dans la recherche en écologie (e.g., Whittaker *et al.*, 1973; Chase et Leibold, 2003; Pocheville, 2015). Il faut attendre les années 2000 pour qu'un cadre théorique réconciliant ces deux aspects du concept soit proposé (Chase et Leibold, 2003; Peterson *et al.*, 2011).

La définition d'une niche proposée par Chase et Leibold (2003) s'appuie à la fois sur l'intégration des prérequis environnementaux de l'espèce, représentés par les variables non interactives (e.g., le climat, la topographie...), et sur les variables interactives, c'est-à-dire les interactions de cette espèce avec l'environnement (e.g., la consommation de ressources). Cette dernière composante est en effet fondamentale puisqu'à une échelle locale, les espèces consomment des ressources et interagissent entre elles. Un tel modèle doit donc représenter un espace environnemental comportant une composante statique (prérequis) et une composante dynamique (interactions). Cela signifie que l'environnement change constamment suivant l'impact de l'espèce sur

lui. Or, l'échelle à laquelle se développe l'étude des distributions biogéographiques par l'ENM ne permet pas de s'ancrer dans un tel cadre théorique (Araújo et Guisan, 2006; Peterson *et al.*, 2011). D'une part, ces modèles de niches sont essentiellement statiques et explorent des corrélations à un instant *t*, ce qui rend les mécanismes de rétroaction difficiles à évaluer (Araújo et Guisan, 2006). D'autre part, ces interactions sont difficiles à mesurer empiriquement ou à détecter pour de grandes échelles de temps et d'espace, puisqu'elles se manifestent de façon très locale avec parfois d'importants changements sur de faibles distances (Araújo et Guisan, 2006; Soberón, 2007; Peterson *et al.*, 2011; Araújo et Rozenfeld, 2013).

Il est possible de contourner ce problème en définissant deux grandes classes de niches en fonction du type de variables utilisées pour les modéliser: la niche Grinnellienne, définie par des variables non-interactives, et la niche Eltonienne, définie par les variables interactives (Soberón, 2007). Ces deux classes de niches ne se manifestent pas à la même échelle géographique, étant donné la résolution à laquelle les deux types de variables sont mesurées. L'estimation des variables non interactives (ou scénopoétiques) concerne des échelles régionales à macro-régionales, voire mondiales, tandis que les variables interactives sont mesurées à une échelle locale (Araújo et Guisan, 2006; Soberón, 2007; Peterson *et al.*, 2011). Dans ce contexte, J. Soberón (2007) propose l'hypothèse du bruit Eltonien (en anglais: « *Eltonian Noise Hypothesis* »). Celle-ci s'appuie sur l'observation de l'effet généralement limité ou peu significatif des interactions biotiques sur la distribution d'une espèce à de grandes échelles spatiales, et propose de ne pas les prendre en compte dans la reconstitution de niches dans le cadre de questionnements sur la distribution macro-géographique d'un taxon ou population. C'est ce point de vue qui est adopté dans le cadre théorique que je présente ici, puisque celui-ci s'intéresse à des phénomènes d'ordre macro-géographique (en référence à Peterson *et al.*, 2011).

*Le concept Hutchinsonien de niche écologique*

L'une des principales avancées dans la définition de niche écologique est le fruit du travail de l'écologue anglais G. Evelyn Hutchinson qui a proposé dans les années 1950 une approche plus mathématique du concept (Hutchinson, 1957). Il définit une niche

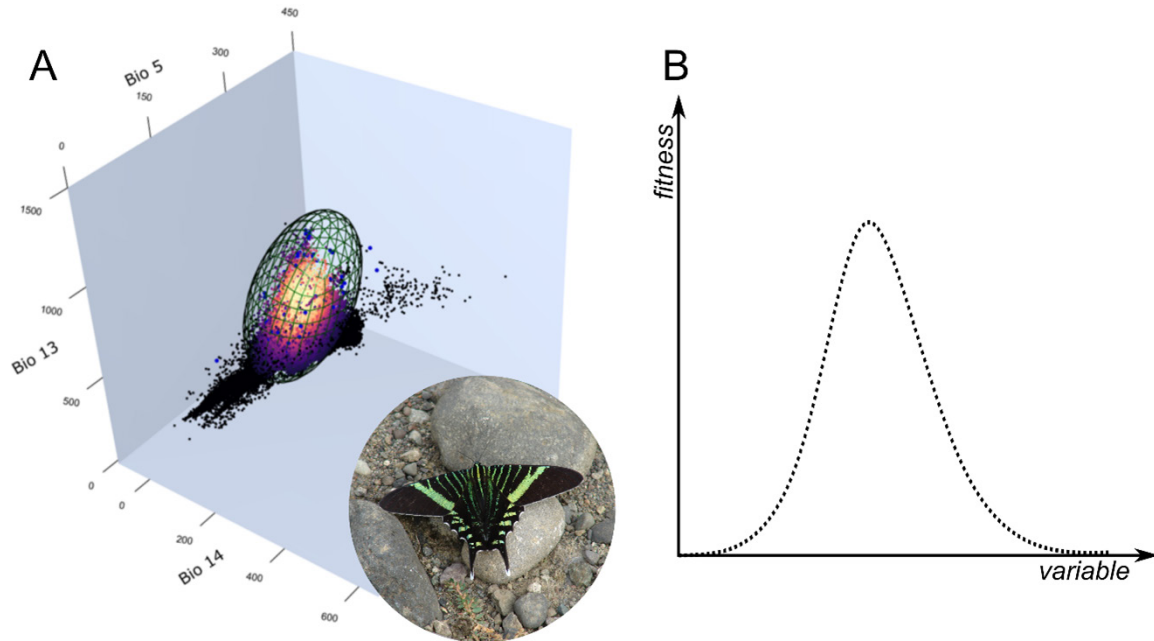


Figure 1. **A.** Modèle de la niche du phalène *Urania fulgens* (occurrence 530489 observée le 23 mars 2011<sup>1</sup>) sous la forme d'un ellipsoïde dans un espace environnemental tri-dimensionnel composé de variables bioclimatiques significativement corrélées à sa distribution géographique (Nuñez-Penichet *et al.*, 2021b). **B.** Représentation schématique de la tolérance physiologique unimodale d'un individu à une variable environnementale.

comme « un hypervolume de  $n$ -dimensions [...], au sein duquel chaque point correspond à un état de l'environnement qui permettrait à l'espèce [...] d'exister indéfiniment. » (*ibid.* p. 416). Elle s'exprime dans l'espace  $n$ -dimensionnel des niches  $N$ , dont chaque dimension correspond à une variable environnementale. Cette définition correspond à la *niche fondamentale* ( $N_F$ ), qui, en théorie, définit les propriétés écologiques intrinsèques de l'espèce considérée en l'absence d'interactions, à partir de variables scénopoétiques (également appelées abiotiques; Figure 1). Ces dernières sont typiquement les variables climatiques ou géographiques (Peterson *et al.*, 2011). G. E. Hutchinson précise également que cet hypervolume ne serait pas une façon de « binariser » l'espace environnemental entre les points où l'espèce a la même probabilité d'exister, qui matérialiseraient  $N_F$  (valeur sélective ou *fitness* = 1), et les points où l'espèce n'a aucune probabilité d'exister, et qui sont donc en dehors de  $N_F$  (*fitness* = 0). Il faudrait plutôt penser la niche comme un gradient de la valeur sélective d'une espèce au sein de  $N$ . Autrement dit, la population considérée a une plus grande probabilité de persistance dans certaines zones de sa niche que dans d'autres (Hutchinson, 1957). L'hypothèse la plus communément admise de nos jours est que la niche fondamentale prend la

forme d'un objet convexe, comme un polyèdre ou un ellipsoïde (*e.g.*, Van Aelst et Rousseeuw, 2009; Escobar *et al.*, 2014, 2017; Qiao *et al.*, 2016; Jiménez *et al.*, 2019; Soberón et Peterson, 2020; Nuñez-Penichet *et al.*, 2021b; Figure 1A). Cette supposition repose sur l'idée que le bord de  $N_F$  doit être de forme convexe, car la tolérance physiologique d'un individu à une variable environnementale est considérée comme étant unimodale par nature (Angilletta, 2009; Drake, 2015; Jiménez *et al.*, 2019; Soberón et Peterson, 2020; Figure 1B). Cependant, certains travaux remettent en partie en cause cette hypothèse, notamment dans le cas d'espèces annuelles vivant dans un environnement fortement marqué par la saisonnalité ou dans le cas d'espèces migratrices (Nakazawa *et al.*, 2004; Soberón et Peterson, 2020; Ingenloff 2020). En effet, un organisme peut « changer » de niche, par exemple entre sa forme juvénile et sa forme adulte (Grubb, 1977; Soberón et Arroyo-Peña, 2017) ou dans le cadre d'un cycle saisonnier/migratoire (*e.g.*, Soberón et Peterson, 2020; Ingenloff 2020). Il faut donc considérer un modèle de niche comme un modèle statique valable à un instant  $t$  (Hutchinson, 1957, p. 417). Cela implique un certain nombre de précautions pour le choix des données à l'origine de la modélisation.

En s'appuyant principalement sur les travaux de V. Volterra (1926) et G.F. Gause (1934), G. E.

<sup>2</sup> source photo: [www.butterfliesandmoths.org](http://www.butterfliesandmoths.org).



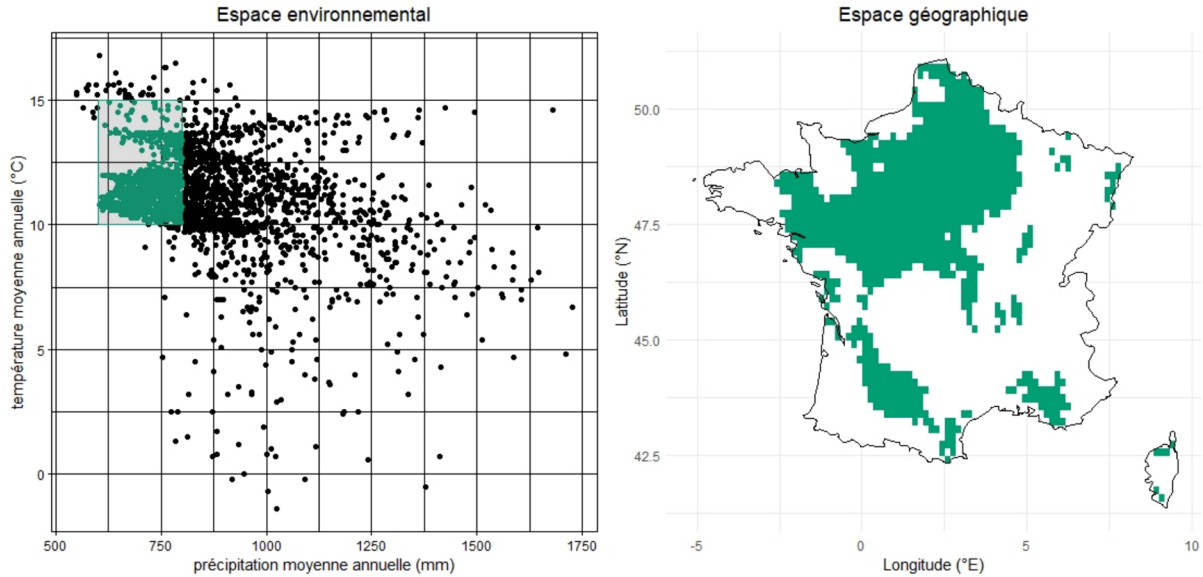


Figure 2. Illustration de la dualité de Hutchinson dans le cas d’une espèce virtuelle et d’un espace environnemental simplifié, dont les deux variables sont la température moyenne annuelle et la précipitation moyenne annuelle. Dans l’espace environnemental, le nuage de points représente les combinaisons de températures et précipitations disponibles en France de nos jours (1979-2013), d’après les simulations issues du modèle CHELSA (résolution spatiale: 10 arc min.; Karger *et al.*, 2017). Le rectangle vert représente la niche de l’espèce virtuelle. Les localités correspondant à ces conditions favorables sont représentées dans l’espace géographique sous la forme de pixels verts. Il est intéressant de noter ici que certaines conditions environnementales incluses dans la niche n’existent pas en France actuellement (zones grises contenue dans le rectangle vert).

Hutchinson inclut dans son modèle une composante biotique qui rend compte de la possibilité de compétition entre deux espèces pour les mêmes portions de l’espace environnemental à une localité donnée. Il formule donc le principe de Volterra-Gause, selon lequel l’occurrence de deux espèces au sein d’une même localité signifie qu’elles doivent forcément occuper des niches écologiques différentes, sans quoi il y a compétition et, *in fine*, disparition de l’une des deux espèces à la localité considérée. Ce principe implique la définition de la *niche réalisée* ( $N_R$ ), qui, hypothétiquement, est une réduction de  $N_F$  par les interactions de compétition avec d’autres espèces (Hutchinson, 1957).

Si l’on considère l’espace environnemental ( $E$ ), il est possible d’identifier des régions correspondant aux conditions environnementales présente dans une ou plusieurs localités de l’espace géographique ( $G$ )<sup>3</sup>, notées  $\eta(G_i)$ . (Figure 2; Hutchinson, 1957; Peterson et Soberón, 2012). À l’inverse, chaque localité de l’espace géographique ( $G$ ) ne correspond qu’à un point de l’espace environnemental ( $E$ ). Ces environnements présents dans  $G$  sont notés  $\eta^{-1}(E)$ . Ce principe est appelé la dualité de Hutchinson, et

<sup>3</sup> Bien que G. E. Hutchinson utilise la notation  $B$  et  $N$  pour désigner le biotope et l’espace des niches, j’emploierai plutôt les abréviations  $G$  pour espace géographique et  $E$  pour espace environnemental dans la suite de ce texte, en référence à un article récent de clarification théorique (Peterson et Soberón, 2012).

a des conséquences théoriques et pratiques cruciales (Colwell et Rangel, 2009). En effet, il implique que la niche s’exprime à la fois dans l’espace environnemental et dans l’espace géographique, et donc que ces deux espaces sont intimement liés. Cette relation n’est toutefois pas réciproque. De façon logique, à une localité de ( $G$ ) ne peut correspondre qu’une seule combinaison de variables environnementales dans ( $E$ ), puisqu’elle est unique dans  $G$ . En revanche, une combinaison de variables environnementales dans ( $E$ ) peut correspondre à plusieurs localités de ( $G$ ); autrement dit, les mêmes conditions environnementales peuvent être présentes à différentes localités géographiques. La conséquence de cette relation non-réciproque est qu’un hypervolume continu dans  $N$  peut correspondre à plusieurs zones discontinues dans  $G$ , et vice-versa (Figure 2). C’est pourquoi il est primordial de prendre en compte ces deux espaces à la fois,  $G$  et  $E$ , lorsque l’on cherche à aborder la question des distributions biogéographiques en relation avec les niches écologiques (Peterson *et al.*, 2011).

La relation entre les deux types de niches de  $G$ . E. Hutchinson peut s’écrire en termes mathématiques selon l’inégalité suivante (Soberón et Arroyo-Peña, 2017):

$$N_F \supseteq N_R$$

Cette inégalité prédit que les combinaisons de variables environnementales correspondant aux localités où l'espèce est présente ( $N_R$ ) sont contenues dans l'hypervolume de  $N_F$ , ce qui se vérifie, ou pas, empiriquement (voir discussion Soberón et Arroyo-Peña, 2017).

Or, il est possible que des combinaisons de variables présentes dans  $N_F$  n'existent pas dans une zone et à un instant donné (Hutchinson, 1957; Pulliam, 2000; Jackson et Overpeck, 2000; Soberón et Peterson, 2011; Figure 2). C'est pourquoi Jackson et Overpeck (2000) ont formalisé l'idée de « niche potentielle », aussi nommée « niche disponible » (Green, 1971) ou « *niche fondamentale existante* », annotée  $N_F^*$  (Peterson *et al.*, 2011, Peterson et Soberón, 2012). Elle se définit par l'intersection entre  $N_F$  et l'espace environnemental disponible à l'espèce considérée, soit en termes mathématiques :

$$N_F^* = N_F \cap E$$

En d'autres termes,  $N_F^*$  correspond à l'ensemble des combinaisons de variables qui existent dans la région et la période considérée et qui coïncident avec les exigences écologiques de l'espèce. Ce troisième type de niche peut être ajouté à l'inégalité de Hutchinson de la façon suivante :

$$N_F \supseteq N_F^*(t) = N_F \cap \eta(M, t) \supseteq N_R(t)$$

Cette inégalité établit un lien explicite entre niche écologique et répartition géographique puisque, par définition,  $N_F$  est avant tout définie dans l'espace environnemental, alors que  $N_F^*$  et  $N_R$  sont définies à partir des occurrences géographiques de l'espèce considérée. Il convient maintenant de décrire plus précisément cette relation entre niches écologiques et répartitions géographiques.

#### Relation entre niche et distribution

Considérons le diagramme **BAM** (Biotique, Abiotique, Mobilité) proposé par Soberón et Peterson (2005, 2012), qui délimite un cadre théorique pour réfléchir aux facteurs influençant la distribution géographique d'un taxon ou une population (Figure 3). La projection de  $N_F$  dans l'espace géographique **G**, c'est-à-dire les localités géographiques présentant des conditions environnementales incluses dans  $N_F$ , identifie les endroits où le taxon pourrait rencontrer des conditions scénopoétiques favorables (**A**). La

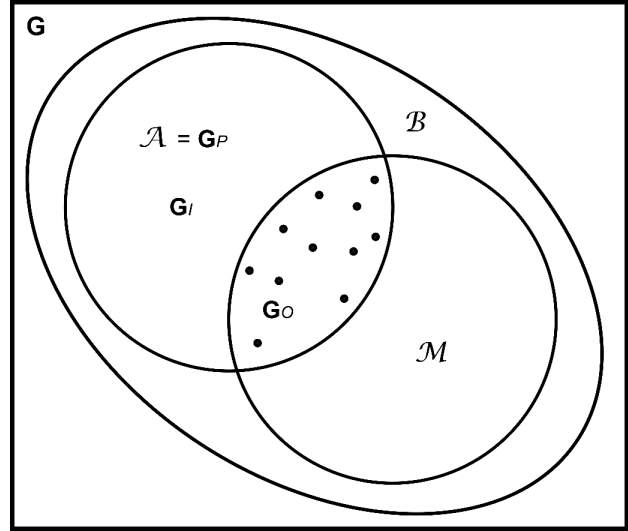


Figure 3. Diagramme **BAM** illustrant les facteurs influençant la distribution macro-géographique d'une espèce dans le cas de l'hypothèse du bruit Eltonien (d'après Soberón et Peterson, 2005, 2012, modifié). Les cercles représentent les différents facteurs et les points noirs représentent la distribution géographique de l'espèce. **G**: espace géographique; **A**: variables scénopoétiques; **B**: interactions biotiques; **M**: zones géographiques accessibles à l'espèce;  $G_p$ : aire de distribution potentielle;  $G_o$ : aire de distribution occupée;  $G_i$ : aire de distribution potentiellement envahissable.

distribution géographique du taxon peut également être réduite par les interactions biotiques (**B**). Cependant, si nous nous plaçons dans l'hypothèse du bruit Eltonien, ce facteur est peu limitant. Enfin, le dernier facteur est l'accessibilité potentielle du taxon à ces conditions favorables sur une période pertinente (**M**; Pulliam, 2000). Ces dernières peuvent en effet être trop éloignées de l'aire de répartition de la population source ou séparées de celle-ci par une barrière géographique infranchissable, comme un océan ou une chaîne de montagne. L'intersection de **A** et **B** est l'aire de distribution potentielle ( $G_p$ ), qui est l'expression géographique de  $N_F^*$ . L'intersection de  $G_p$  avec **M** définit l'aire de distribution occupée ( $G_o$ ), qui est l'expression géographique de  $N_R$ . Enfin, la zone de  $G_p$ , qui n'est pas occupée, est appelée l'aire de distribution envahissable ( $G_i$ ), c'est-à-dire que les conditions sont favorables et que le taxon n'y pas encore accès, mais qu'il pourrait théoriquement y perdurer.

Dans ce cadre, il est possible d'établir au moins quatre configurations du diagramme **BAM** (Peterson *et al.*, 2011; Saupe *et al.*, 2012; Figure 4): i) la configuration classique, dans laquelle **A** et **M** se recoupent partiellement; ii) la configuration de Hutchinson, dans laquelle la présence de conditions

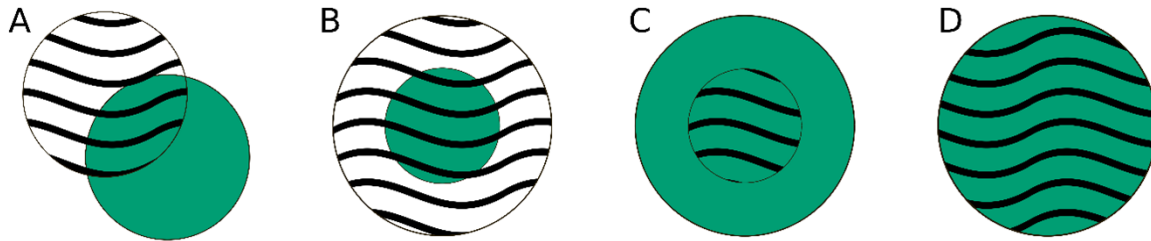


Figure 4. Quatre scénarios théoriques du diagramme **BAM** (d'après Saupé *et al.*, 2012, modifié). Le cercle vert représente les conditions scénopécologiques favorables (**A**); le cercle aux hachures ondulées représente l'aire accessible à l'espèce (**M**). **A.** Configuration classique. **B.** Configuration de Hutchinson. **C.** Configuration de Wallace. **D.** Configuration de superposition.

favorables constitue le principal facteur limitant à l'établissement de l'espèce dans un endroit et à un moment donnés, soit **M** englobe **A**; iii) la configuration de Wallace, dans laquelle le principal facteur limitant est l'accès aux conditions favorables, soit **A** englobe **M**; et enfin, iv) la configuration de superposition, dans laquelle **A** et **M** se superposent parfaitement. Le choix d'une configuration parmi ces quatre-là peut avoir un impact significatif sur le résultat de la modélisation de la niche (Saupé *et al.*, 2012). En effet, une étude a montré que les modèles sont plus performants dans la configuration classique et dans celle de Hutchinson (*ibid.*); il conviendra donc de se placer dans l'une ou l'autre de ces hypothèses.

Il faut noter ici l'importance que prend **M** comme facteur contraignant de la distribution géographique d'un taxon ou d'une population, en plus de la présence de conditions favorables. Il sera donc capital d'en donner une estimation fiable pour la bonne conduite d'une modélisation de niche. Elle doit correspondre à une réalité biogéographique plutôt qu'à des frontières administratives; ses limites dépendent des capacités dispersives du taxon et sont par exemple matérialisées par des barrières géographiques naturelles, telles que la présence d'une chaîne de montagne, d'un océan ou d'une rivière (Pulliam, 2000; Barve *et al.*, 2011; Machado-Stredel *et al.*, 2021).

#### *Que modélise-t-on en ENM ?*

La modélisation de niches écologiques grinnelliennes s'appuie sur deux grands types de données: d'une part, les localités géographiques où le taxon considéré a été observé, et d'autre part, les conditions environnementales présentes dans l'aire géographique accessible au taxon. À partir de ces données, l'objectif est de déterminer, dans chaque pixel de la région étudiée, la probabilité d'occurrence en fonction des conditions environnementales présentes dans ce pixel. En d'autres termes, nous

créons un modèle pour la fonction qui décrit la relation entre les variables environnementales et les occurrences du taxon (*i.e.*, la niche; Peterson *et al.*, 2011). Pour cela, nous employons un ou des algorithmes prédictifs qui établissent des règles décrivant la relation entre différentes variables ou paramètres, puis qui utilisent ces règles pour reconstruire la niche. Afin d'interpréter correctement les modèles résultants, il est primordial de déterminer le ou les types de niches –  $N_F$ ,  $N_F^*$  ou  $N_R$  – qui peuvent potentiellement correspondre à ces données de sortie.

La nature des données utilisées (données d'occurrences et variables environnementales) réduit d'ores et déjà la possibilité de modéliser tous les types de niches présentés précédemment. En effet, ces données sont avant tout géographiques. Or, nous avons vu que la niche fondamentale est définie en premier lieu dans l'espace environnemental, et que dans l'espace géographique, elle peut être réduite par au moins deux types de facteurs: les interactions avec d'autres espèces et l'aire géographique accessible au taxon. Cela signifie que la niche fondamentale ne peut être estimée à partir de données géographiques seulement, étant donné que la distribution géographique du taxon ne correspond pas nécessairement à **A** (Jiménez *et al.*, 2019). Elle doit être avant tout définie à partir de données expérimentales visant à déterminer les conditions limites tolérées par les individus (Soberón et Arroyo-Peña, 2017). La niche modélisée dans le cadre de l'ENM se situe plus probablement quelque part entre la niche fondamentale existante et la niche réalisée en fonction de l'algorithme et des données utilisés.

La nature de la démarche conditionne également l'interprétation des cartes de répartition issues de ces modélisations. La modélisation de distribution d'espèces (*Species Distribution Modeling*, abrégé. SDM) peut parfois employer les mêmes jeux de données et algorithmes qu'en ENM (Peterson et

Soberón, 2012). Toutefois, le cadre théorique et les concepts sous-jacent à la SDM sont différents et n'impliquent pas les mêmes axiomes – en particulier dans l'interprétation des cartes de répartition. En SDM, les cartes de sortie sont interprétées comme la distribution potentielle d'une espèce, tandis qu'en ENM, elles correspondent à des projections géographiques d'une niche définie dans l'espace environnemental à partir de la répartition géographique connue d'une espèce (*ibid.*; Peterson *et al.*, 2011; Sillero, 2011; Araújo et Peterson, 2012; Warren, 2012; Sillero *et al.*, 2021). Cette différence interprétative a des conséquences sur le choix des données d'occurrences et des algorithmes; ainsi, les études de SDM utilisent généralement des données de présence-absence, tandis qu'en ENM, les données de présence-seule et de présence-*background* sont préférées (*ibid.*; cf. 3.2.). La qualité d'une étude employant l'ENM repose donc avant tout sur l'exposé clair et explicite de ses fondements conceptuels (Peterson et Soberón, 2012). Ceux-ci permettent de guider le choix des données à l'origine des modèles ainsi que la conception des analyses, pour que les modèles finaux se rapprochent au maximum des phénomènes que l'on souhaite étudier, qu'ils soient écologiques ou culturels. Dans la suite de cet article, je propose un certain nombre de recommandations quant au choix des données et des algorithmes en ENM. Je me focaliserai donc principalement sur les méthodes de présence-seule ou présence-*background*, et plus particulièrement sur Maxent, l'algorithme le plus populaire en ENM.

#### CHOISIR LES DONNEES A L'ORIGINE DES MODELES DE NICHES ECOLOGIQUES: QUELQUES PRECAUTIONS METHODOLOGIQUES

##### *Données environnementales*

Les données environnementales utilisées en ENM sont des cartes raster (*i.e.*, une carte de données spatiales organisées sous la forme de pixels. À chaque pixel est attribué un set de coordonnées spatiales et une valeur attributaire) de différentes variables pertinentes pour décrire l'environnement du taxon ou de la population considéré. Le choix des variables environnementales comporte trois aspects essentiels.

En premier lieu, il est important de sélectionner des variables abiotiques susceptibles d'influencer la répartition géographique de l'espèce ou population considérée (Peterson *et al.*, 2011), une nécessité qui découle du diagramme **BAM** lorsque l'on se

place dans l'hypothèse du bruit Eltonien (Figure 3). En général, il s'agit de variables climatiques et géographiques (Pearson et Dawson, 2003), mais d'autres types peuvent être employés en fonction de l'espèce étudiée. Par exemple, il est particulièrement pertinent d'inclure des variables décrivant les propriétés du sol pour modéliser la niche écologique de végétaux (*e.g.*, Hengl *et al.*, 2017; Zuquim *et al.*, 2020), ou encore les propriétés de l'eau dans le cas d'espèces aquatiques (*e.g.*, Domisch *et al.*, 2015; Sbrocco et Barber, 2013). En revanche, l'élévation n'est généralement pas une variable pertinente en ENM en raison de son importante corrélation avec la température. L'utilisation de variables corrélées n'a pas nécessairement de conséquences majeures sur les modèles (en tout cas, avec l'algorithme Maxent; Feng *et al.*, 2019b). Cependant, la corrélation entre température et élévation n'est pas constante et diffère dans l'espace (notamment par rapport à la latitude) et dans le temps (Peterson *et al.*, 2011, p. 85). Cela est problématique lorsque l'on cherche à transférer le modèle à une autre région ou une autre période. Il est plutôt recommandé d'utiliser des indicateurs topographiques calculés à partir de l'élévation, mais qui sont indépendants des variables climatiques (*e.g.*, Amatulli *et al.*, 2018, 2019).

Dans tous les cas, la plupart de ces variables correspondent elles-mêmes à des modèles, étant donné qu'il n'est pas possible de les mesurer en continu sur d'aussi grandes surfaces (Figure 5), et ce particulièrement concernant des variables paléoclimatiques qui reconstituent le climat et l'environnement pour le passé (*e.g.*, Valdes *et al.*, 2017; Boucher *et al.*, 2020), pour lequel la densité spatio-temporelle des proxys climatiques est très faible. En conséquence, les variables basées sur des modèles peuvent comporter un certain nombre d'incertitudes dont il faut être conscient lors de leur emploi en ENM (Varela *et al.*, 2015). Il est alors important de consulter leurs métadonnées afin de vérifier la fiabilité globale de ces modèles, mais également leur fiabilité à l'échelle de la région d'intérêt. Par exemple, pour des jeux de variables climatiques mondiales telles que WorldClim (Fick et Hijmans, 2017), certaines zones sont plus fiables que d'autres en raison de l'inégale densité de répartition des stations météorologiques (Figure 5). Dans le cas de variables paléoclimatiques, certaines périodes sont également mieux documentées que d'autres, ce qui permet une meilleure reconstitution du climat (*e.g.*,



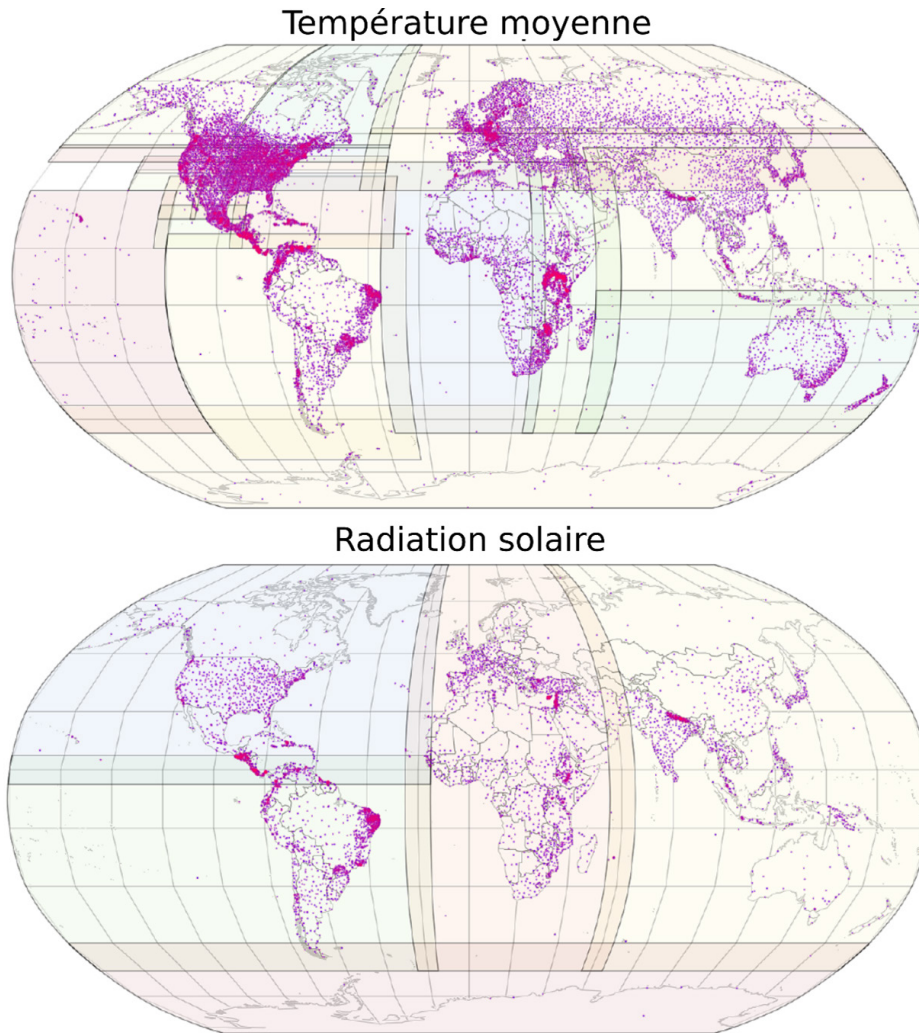


Figure 5. Localisation des stations météorologiques utilisées pour produire les variables climatiques par interpolation, exemple de la température moyenne et de la radiation solaire du jeu de données WorldClim2 (Fick et Hijmans, 2017). L'inégale densité de stations météorologiques implique que la reconstitution des variables sera plus précise dans certaines régions que d'autres.

le dernier maximum glaciaire en Europe; Varela *et al.*, 2015).

Ensuite, il est important de prendre en compte la dimensionnalité de l'environnement modélisé, c'est-à-dire le nombre de variables employées. Il est préconisé d'utiliser un nombre restreint de variables non ou peu corrélées. L'utilisation d'une grande quantité de variables rend le modèle plus complexe et peut entraîner une trop grande adéquation entre le modèle de niche et les données d'occurrences – *overfitting* en anglais (Peterson *et al.*, 2011, p. 87) – ce qui peut masquer certains aspects écologiques de la niche modélisée. La trop grande complexité d'un modèle de niche diminue la qualité des modèles (*e.g.*, Warren et Seifert, 2011), et donc les possibilités d'effectuer des analyses postérieures, tel que le transfert à une autre région

ou une autre période (Peterson et Nakazawa, 2007). Il est en ce sens primordial que le nombre de variables environnementales soit inférieur au nombre d'occurrences (Sillero *et al.*, 2021). En outre, l'utilisation de variables hautement corrélées entraîne l'impossibilité d'utiliser certains types d'algorithmes (Guisan et Zimmerman, 2000). Enfin techniquement, utiliser un trop grand nombre de variables augmente automatiquement le temps de calcul et de calibration des modèles, une contrainte purement matérielle mais non négligeable dans la conduite d'un projet inscrit dans un temps limité. Il existe de nombreuses propositions afin de sélectionner les meilleures variables pour un modèle de niche donné, en fonction des objectifs de l'étude (voir Cobos *et al.*, 2019a pour une revue de la littérature).

Enfin, le dernier élément à prendre en compte est la résolution spatiale et l'étendue chronologique des variables employées. Celles-ci doivent impérativement correspondre à la résolution des données d'occurrences, au risque de produire un modèle de niche incohérent d'un point de vue spatio-temporel (Peterson *et al.*, 2011, p. 91; Sillero et Barbosa, 2020). Si les cartes de données environnementales présentent une résolution spatiale trop grossière (*i.e.*, des pixels trop larges) par rapport à l'échelle à laquelle se manifeste le phénomène étudié, il sera nécessaire de réaliser une descente d'échelle statistique afin d'augmenter la résolution. Cette étape est assez délicate, car simplement subdiviser les pixels afin d'obtenir un maillage plus fin n'augmente pas la résolution intrinsèque de l'information donnée par la variable en question (Sillero et Barbosa, 2020). Il est nécessaire pour cela d'avoir « *accès à la relation d'interpolation utilisée pour [...] estimer [ladite variable] et un modèle d'élévation à la résolution adéquate* » (*ibid.*, p. 5). Plusieurs méthodes existent, des plus simples – comme l'interpolation avec le plus proche voisin – aux plus complexes – comme la descente d'échelle par modélisation GAM (*Generalized Additive Model*; *e.g.*, Vrac *et al.*, 2007; Antunes, 2015; Latombe *et al.*, 2018) ou la méthode de correction delta (*e.g.*, Beyer *et al.*, 2020). D'autre part, la correspondance chronologique entre les données d'occurrences et les données environnementales est essentielle, au risque d'associer les mauvaises combinaisons de valeurs aux occurrences recensées. Par exemple, dans le cas d'espèces migratrices, l'emploi d'un jeu de données trop imprécis temporellement peut conduire à des modèles peu informatifs sur leur comportement écologique (Ingenloff, 2020; Ingenloff et Peterson, 2021).

Plusieurs auteurs ont souligné l'impact du choix des variables environnementales sur les modèles de niches sous-jacents (*e.g.*, Peterson et Nakazawa, 2007; Diniz-Filho *et al.*, 2009; Varela *et al.*, 2015; Cobos *et al.*, 2019a, b; Vignoles, 2021). En effet, en ce qui concerne les variables climatiques, de nombreux modèles du climat régional ou global existent, plus ou moins complexes, ne se basant pas sur les mêmes types de calculs ni les mêmes processus influençant le climat (*e.g.*, Modèles de Circulation Globale; Kageyama *et al.*, 2005; Singarayer et Valdes, 2010; Valdes *et al.*, 2017;

Boucher *et al.*, 2020; Modèles de Système Terrestre de Complexité Intermédiaire; Claussen *et al.*, 2002; Goosse *et al.*, 2010; Roche *et al.*, 2014). Chaque modèle fournit donc des jeux de simulations plus ou moins différents, influençant de fait les modèles de niches sur lesquels ils seront basés (Varela *et al.*, 2015). Ce facteur de variabilité est donc important à prendre en compte, en particulier lorsque l'étude se propose d'évaluer l'impact du changement climatique sur la répartition géographique des aires environnementales en adéquation avec les nécessités écologiques d'un taxon ou d'une population (*e.g.*, Diniz-Filho *et al.*, 2009; Peterson *et al.*, 2018). Il est donc préconisé de toujours proposer une évaluation de la variance des modèles en fonction des différentes sources d'incertitudes (*e.g.*, Diniz-Filho *et al.*, 2009; Cobos *et al.*, 2019b); il est également intéressant de présenter la répartition spatiale de la variance ou de l'écart-type, afin d'identifier les régions dans lesquelles la variabilité est plus forte ou, au contraire, plus faible (*e.g.*, Peterson *et al.*, 2018; Cobos *et al.*, 2019b; Warren *et al.*, 2021a).

#### *Données d'occurrences*

Les données d'occurrences permettent de rendre compte de la répartition géographique de l'espèce ou de la population étudiée. Elles se caractérisent par l'observation de la présence ou de l'absence d'un taxon dans une localité géographique donnée et à un moment donné.

Cette observation est soumise à plusieurs types de facteurs dont il faut être conscient lorsque l'on tente d'estimer la répartition géographique d'une espèce (Peterson *et al.*, 2011, p. 63). Les facteurs qui intéressent les études d'ENM sont les facteurs biologiques (Figure 6), qui concernent donc la valeur sélective de l'espèce ou de la population étudiée – en d'autres termes, il s'agit de la  $N_{f^2}$  dont la projection géographique est **A** dans le diagramme **BAM** modifié (Figure 3). Cette répartition géographique est également conditionnée par l'accessibilité de l'espèce ou population à **A** (**M**). Enfin, le dernier type de facteurs est plus problématique, car il introduit des biais qui ne dépendent pas de l'écologie de l'espèce considérée, mais plutôt du fonctionnement de la recherche (Figure 6). Il s'agit par exemple de la différence de détectabilité entre espèces (*e.g.*, Royle et Dorazio, 2006) – certaines espèces sont tout simplement plus faciles à détecter que d'autres – ou encore la disparité dans les efforts de collecte de

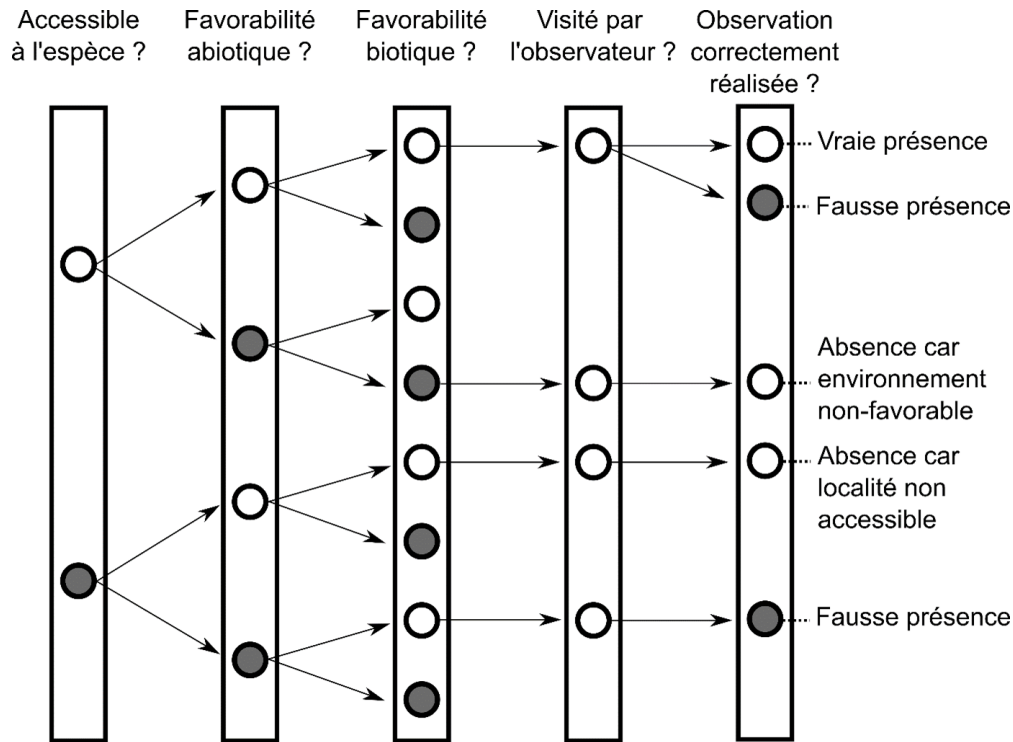


Figure 6. Événements probabilistes menant à l'observation d'une présence ou d'une absence, en incluant des résultats erronés. Chaque barre représente un choix, les cercles blancs représentent « oui » et les cercles gris représentent « non ». L'observation d'une présence ne requiert pas seulement que l'espèce soit présente dans une localité en fonction des trois processus biologiques représentés dans les trois premières colonnes, mais également que la localité ait été visitée par des observateurs et que cette observation ait été réalisée correctement (ex. pas d'erreur d'identification ou erreur typographique). Plusieurs opportunités existent menant à des erreurs. Dans cet exemple sont soulignées une fausse absence due à une mauvaise observation, deux absences liées à deux causes biologiques radicalement différentes, et une fausse présence résultant d'une mauvaise observation ou enregistrement (Figure et légende issues de Peterson *et al.*, 2011, fig. 5.1, modifiées). L'analyse probabiliste de cet arbre par les auteurs de la Figure conclut que la part de présence dans l'estimation de la répartition géographique d'une espèce est plus fiable que la part d'absence, puisque la probabilité d'identifier une fausse présence est moins importante que celle d'identifier une fausse absence.

certains taxons ou dans certaines régions (Peterson *et al.*, 2011, p. 70-71). Un biais spatial courant est par exemple la concentration d'occurrences le long de routes, car il est statistiquement plus fréquent d'y recenser une occurrence que dans des zones moins fréquentées (Kadmond *et al.*, 2004). Un autre biais possible est la qualité du recensement de l'occurrence : l'identification d'un taxon peut comporter des erreurs qui dépendent de l'expérience du collecteur ou de l'état d'avancement de la recherche (localité recensée avant la découverte dudit taxon). Dans le cas de données paléontologiques ou archéologiques, un biais important est également la conservation différentielle des vestiges liée aux conditions d'enfouissement et de fossilisation. Celles-ci conduisent à un biais d'échantillonnage drastique s'ajoutant aux autres biais susmentionnés, car seule une très faible minorité de vestiges rencontrent des conditions suffisantes pour se conserver dans le sol.

Ce biais est d'autant plus fort que l'étendue spatio-temporelle du taxon ou population étudié est faible : plus un taxon a une répartition géographique et une durée de vie limitées, moins il a de probabilité d'être échantillonné (Signor et Lipps, 1982).

Ces quelques exemples montrent que les données d'occurrences ne peuvent être interprétées simplement comme la documentation de l'absence et présence d'un taxon. Elles résultent en réalité d'une complexe intrication de facteurs écologiques (qui nous intéressent) et de facteurs extérieurs (dont il faut s'affranchir au maximum ; Figure 6). Ces derniers auront pour conséquence d'introduire des biais spatiaux dans l'estimation de la répartition géographique de l'espèce, qui seront plus ou moins limitant pour la suite de l'étude. Ceux-ci ne se répercutent pas nécessairement sur l'échantillonnage des environnements occupés par le taxon étudié (Peterson *et al.*, 2011). La présence plus ou moins

importante de ces biais conditionne également le type d'algorithme qui sera employé par la suite. En particulier, l'addition de biais d'échantillonnage successifs aux facteurs écologiques conduit généralement à une estimation plus fiable de la part de présence (avec moins de probabilités d'aboutir à une fausse présence) que de la part d'absence (plus de probabilité d'aboutir à une fausse absence; *ibid.*; Figure 6). Les jeux d'occurrences marqués par de tels biais devront donc être employés avec des algorithmes adaptés, dits de présence-seule ou de présence-*background*. En ENM, les données d'absence sont en effet rarement employées, car les méthodes de présence-absence ont plutôt pour objectif de « *distinguer les conditions environnementales entre habitats occupés et non occupés, résultant en la probabilité de rencontrer l'espèce à chaque localité* » (Sillero *et al.*, 2021, p. 4; Sillero, 2011). Cela découle de la moins bonne fiabilité des données d'absence par rapport aux données de présence: une donnée d'absence peut en effet correspondre à une localité aux environnements favorables, mais inaccessible à l'espèce (Peterson *et al.*, 2011, p. 76). Les modèles résultants correspondent de ce fait à des cartes de probabilité de présence – un objectif de la SDM plutôt que de l'ENM. (*ibid.*; Sillero *et al.*, 2021). En ENM, il est préférable de n'employer que des modèles utilisant des données de présence-seule ou de présence-*background*.

En outre, les biais d'échantillonnage peuvent se répercuter sur l'évaluation du modèle. En effet, la sur-représentation de l'espèce dans une zone va conduire à ce que les occurrences servant à tester le modèle soient géographiquement et/ou environnementalement trop proches de celles servant à le calibrer. Ces effets spatiaux et/ou environnementaux vont donc conduire à de moins bonnes prédictions. Pour limiter ces biais, le procédé de raréfaction spatiale ou filtration spatiale (en anglais, *spatial thinning* ou *spatial filtering*) est le plus communément utilisé et recommandé (*e.g.*, Anderson et Gonzalez, 2011; Boria *et al.*, 2014). Il s'agit d'éliminer les occurrences trop proches en fonction d'une distance choisie par l'utilisateur. Plusieurs méthodes de raréfaction spatiale existent (*e.g.*, Aiello-Lammens *et al.*, 2015). Toutefois, ce procédé n'est pas toujours suffisant pour s'affranchir de biais d'échantillonnage, en particulier si la région d'étude est très hétérogène d'un point de vue environnemental (Varela *et al.*, 2014). Il est en effet important que cette raréfaction des jeux de données

d'occurrences ne conduise pas à l'élimination de conditions environnementales pertinentes pour la création du modèle, en raison de leur proximité géographique avec d'autres points. Il est alors conseillé de recourir à un filtre environnemental en éliminant les points présentant des combinaisons de valeurs trop proches (*ibid.*). Enfin, la combinaison de données issues de différentes sources (*e.g.*, programmes citoyens de science, inventaires de spécimens conservés dans les musées, collecte systématique dans le cadre d'une étude de terrain) peut se révéler problématique, car chaque source est affectée de biais différents (*e.g.*, Fletcher *et al.*, 2019). Il peut alors s'avérer difficile de limiter ces derniers dans un jeu de données unifié, puisque chaque source n'aura pas les mêmes motifs ni les mêmes sources de biais. Plusieurs protocoles existent toutefois pour tenter d'intégrer différentes sources au sein d'une même étude de façon plus pertinente, par exemple en accordant plus de poids à certaines sources par rapport à d'autres ou encore, en comparant des modèles basés sur différentes sources (*ibid.*).

Une donnée d'occurrence est caractérisée par trois informations: sa localisation géographique, son attribution taxonomique et sa temporalité. Selon les sources employées et la qualité intrinsèque des données d'occurrences, ces informations peuvent comporter plus ou moins d'erreurs ou être plus ou moins précises. Il est donc particulièrement important d'évaluer la fiabilité de ces trois informations pour chaque occurrence, afin d'uniformiser le jeu de données, de s'assurer que chaque type d'information possède le même niveau de précision et, le cas échéant, de supprimer les occurrences dont la qualité est trop faible pour l'étude concernée (*e.g.*, Sillero *et al.*, 2021). Par exemple, la localisation de l'occurrence peut être plus ou moins précise selon la qualité et la source de l'information: elle peut se résumer à une indication vague, comme le nom de la commune ou du pays de collecte, ou à l'inverse être associée à une position géo-référencée par GPS. Dans ce cas, il est par exemple conseillé d'associer à chaque occurrence une notion d'incertitude (Wieczorek *et al.*, 2004; Chapman *et al.*, 2020), puis de ne garder que les occurrences se situant en dessous d'un seuil prédéfini (que l'on peut fixer par exemple à la résolution des données environnementales). Il est également fréquent que les coordonnées de localisation comportent des erreurs, comme une inversion de la latitude/longitude, ou



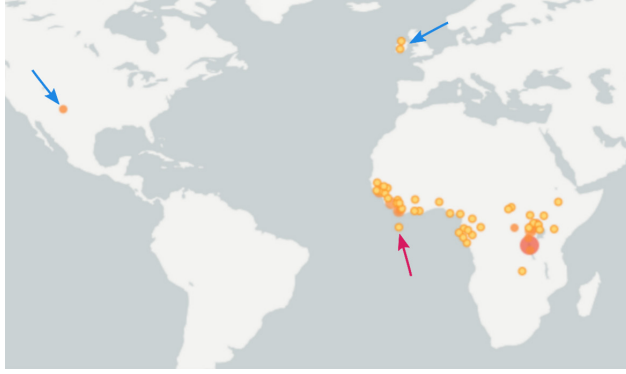


Figure 7. Résultats d'une recherche d'occurrences dans l'agrégateur de données d'occurrences GBIF pour le chimpanzé commun (*Pan troglodytes*) entre 1970 et 2022. La recherche fournit 3 001 occurrences de présence (points jaunes à orange), mais une partie d'entre elles comportent probablement des erreurs ou des incertitudes trop importantes pour être prises en compte dans une étude employant l'ENM. Les erreurs les plus évidentes que nous pouvons relever visuellement sont indiquées par des flèches. La flèche fuchsia indique une occurrence aux coordonnées (0°, 0°), tandis que les flèches bleu clair indiquent des occurrences trop éloignées du reste de l'aire de répartition géographique. Ces quatre localités sont clairement situées dans des environnements défavorables à la survie de *Pan troglodytes* – à savoir l'océan pour la flèche fuchsia et des régions tempérées pour les flèches bleu clair. Le reste du jeu de données nécessitera une vérification plus poussée de la fiabilité des informations taxonomiques, géographiques et temporelles. Source: GBIF.org (24 March 2022) GBIF Occurrence Download<sup>3</sup>

encore une incohérence écologique de la localisation (cas d'occurrences situées dans un océan pour une espèce terrestre ou occurrence isolée très éloignée du reste de la répartition géographique; Sillero *et al.*, 2021; Figure 7). Ces occurrences doivent donc systématiquement être vérifiées, et écartées si les informations erronées ne peuvent être corrigées. Le même type de problème peut apparaître pour les informations de temporalité et d'attribution taxonomique, et nécessiteront également un protocole permettant d'estimer l'incertitude des informations pour sélectionner les données les plus adaptées aux questionnements et à la résolution de l'étude.

À l'issue de cette étape, il est possible de se rendre compte que le taxon étudié ne se prête pas à la problématique de départ. Par exemple, étudier l'évolution d'un groupe au niveau spécifique n'est possible que dans la mesure où suffisamment d'occurrences sont identifiées au niveau de l'espèce. Cet exemple peut sembler caricatural, mais il existe une infinité de cas pour lesquels mesurer l'adéquation des données à la problématique de départ est plus délicat. La construction de ce jeu de données d'occurrences doit donc s'accompagner d'une réflexion sur cette

possibilité d'inadéquation, et potentiellement à une adaptation des questionnements de départ. Une autre question pouvant se poser à ce stade est le nombre d'occurrences qui composent le jeu de données final. Il n'existe pas de règle fixe quant au nombre minimal d'occurrence nécessaires, si ce n'est qu'il doit être le plus élevé possible sans pour autant sacrifier la qualité des occurrences. Une étude a par exemple montré que 3 occurrences peuvent suffire à produire un modèle de niche robuste pour des espèces à faible étendue spatiale, *versus* 13 occurrences pour des espèces plus largement répandues (Prosdij *et al.*, 2016). Toutefois, si trop peu d'occurrences sont utilisées (< 25 localités), la variabilité des modèles et des statistiques d'évaluation augmente (*e.g.*, Hernandez *et al.*, 2006; Wisz *et al.*, 2008). Plusieurs protocoles existent pour limiter les problèmes induits par un trop faible nombre d'occurrences dans la création et l'évaluation des modèles de niches: par exemple, l'évaluation par *jackknife* ou « *leave-one-out* » (Pearson *et al.*, 2006; Shcheglovitova et Anderson, 2013; Gibert, Vignoles *et al.*, 2022) permet d'évaluer la variabilité du modèle en fonction de l'occurrence utilisée pour l'évaluation du modèle. Pour prendre en compte plus de variables environnementales que le nombre d'occurrences, il est aussi possible d'utiliser les « ensembles de petits modèles » (Breiner *et al.*, 2015), qui consistent à combiner plusieurs modèles n'utilisant qu'une petite partie des variables environnementales à chaque fois (*e.g.*, modèles bivariés). Enfin, l'algorithme Maxent semble bien adapté aux petits jeux de données d'occurrences, comme l'ont démontré plusieurs études comparatives (*e.g.*, Hernandez *et al.*, 2006; Papeş et Gaubert, 2007; Wisz *et al.*, 2008)

#### QUELQUES ELEMENTS A PRENDRE EN COMPTE DANS LE PROCESSUS DE MODELISATION *Quels algorithmes pour l'ENM ?*

Il existe de nombreux algorithmes permettant de proposer un modèle de niche écologique, chacun caractérisé par des propriétés intrinsèques qu'il faut connaître afin de le choisir. Le choix d'un algorithme dépend du type de données d'occurrences: de présence-absence (*e.g.*, les forêts d'arbres décisionnels [*random forest*]; Breiman, 2001; les modèles hiérarchiques [*hierarchical models*]; Royle et Dorazio, 2006; les modèles linéaires généralisés [GLM] ou les modèles additifs généralisés [GAM]; Guisan *et al.*, 2002), de présence-seule (*e.g.*, Bioclim;

<sup>3</sup> <https://doi.org/10.15468/dl.rz7r5t>.

Booth *et al.*, 2014; Domain; Carpenter *et al.*, 1993; Habitat; Walker et Cocks, 1991; ou les modèles basés sur la distance de Mahalanobis; Farber et Kadmon, 2003) ou de présence-*background* (e.g., l'algorithme généralisé pour la production de groupes de règles [GARP]; Stockwell et Noble, 1992; ou Maxent; Philips *et al.*, 2006, 2017) et des objectifs de l'étude. Comme souligné par plusieurs auteurs (e.g., Antunes, 2015; Qiao *et al.*, 2015), il n'existe pas d'algorithme objectivement et universellement meilleur qu'un autre en ENM. Leur performance dépend principalement de leur adéquation avec les données et les objectifs de l'étude.

Dans cet article, je propose de nous focaliser sur le fonctionnement de l'algorithme Maxent (Philips *et al.*, 2006, 2017; Philips et Dudik, 2008), en raison de sa popularité dans les études ENM. Celle-ci résulte de sa facilité d'utilisation, rendue possible grâce à un logiciel *stand-alone*<sup>5</sup>, à de nombreux *packages* R spécialisés dans son application et son optimisation (e.g., *kuenm*, *maxnet*, *ENMeval*, *ENMtools*, etc.; Muscarella *et al.*, 2014; Cobos *et al.*, 2019c; Kass *et al.*, 2021; Warren *et al.*, 2021b; Phillips, 2021) et à ses bonnes performances par rapport à d'autres algorithmes (e.g., Phillips *et al.*, 2006; Elith *et al.*, 2006; Hernandez *et al.*, 2006). Cette popularité masque les limites de cet algorithme ainsi que ses subtilités d'application – en particulier la nécessité de dépasser les paramètres par défauts de l'algorithme (Morales *et al.*, 2017). Dans les paragraphes suivants, je présenterai donc plusieurs considérations permettant de correctement l'employer; je propose également une courte section sur une alternative à ce type de modèle, en particulier lorsque l'on veut se focaliser les dynamiques et comportements des niches écologiques dans l'espace environnemental.

*Fonctionnement de l'algorithme Maxent.* « Maxent » est le nom d'un algorithme utilisant le principe d'entropie maximale (Phillips *et al.*, 2006, 2017; Phillips et Dudik, 2008), c'est-à-dire que « la distribution de probabilité estimée doit être en accord avec ce qui est connu [...], mais doit éviter les suppositions qui ne sont pas compatibles avec les données. » (Peterson *et al.*, 2011, p. 109). Il est adapté aux données d'occurrences de type présence-seule, mais emploie des informations sur la variation de l'environnement (dénommé « *background* ») autour des points d'occurrences dans la construction du modèle: c'est donc un modèle de type « présence-*background* ». Concrètement, Maxent va établir

une relation entre l'occurrence d'une présence et la densité des variables environnementales qui lui sont associées, tout en prenant en compte la densité des variables environnementales associées à des pixels sélectionnés aléatoirement dans la zone étudiée (points de *background*). Il en résulte une distribution de probabilité, qui fait ensuite l'objet d'une transformation logarithmique pour représenter la probabilité de présence de conditions adéquates (en anglais, « *suitability* », que je traduirai dans cet article par « favorabilité »; *ibid.*, p. 109), soit des conditions proches des valeurs moyennes observées pour les points d'occurrences (Elith *et al.*, 2011). En termes de statistiques, Maxent va estimer le ratio entre la densité conditionnelle des variables environnementales au niveau des points de présence  $f_1(\mathbf{z})$ , et la densité marginale (i.e., non conditionnelle) des variables environnementales au niveau des points de *background*  $f(\mathbf{z})$ . Parmi toutes les formules de  $f_1(\mathbf{z})$  possibles, Maxent choisit celle qui est la plus proche de  $f(\mathbf{z})$ , selon le principe d'entropie maximale (*ibid.*, p. 47). *In fine*, la fonction doit coller au maximum avec les données de présence, tout en apportant le moins d'informations possible sur le reste de l'aire prise en compte.

En réalité, Maxent calibre le modèle final non pas à partir des variables environnementales brutes, mais à partir de transformations de ces variables, appelées fonctions (*features*). Cette fonctionnalité permet à Maxent de modéliser des relations complexes entre les données de présence et la densité des variables environnementales. Il existe cinq classes de *features*: *linear*, *quadratic*, *product*, *threshold* et *hinge* (Phillips *et al.*, 2006, 2017). Une *linear feature* correspond à la variable continue en elle-même. Le carré de cette variable constitue la *quadratic feature*. *Product* est le produit de deux variables. *Threshold* consiste à attribuer la valeur de 1 si  $f$  est au-dessus d'une certaine valeur, ou 0. *Hinge* est similaire à *threshold*, mais utilise une fonction linéaire au lieu d'une fonction en escalier. Le choix des types et du nombre de *features* utilisés nécessite un travail de réglage par l'utilisateur car les paramètres par défaut sont souvent inadaptés (e.g., Anderson et Gonzalez, 2011; Elith *et al.*, 2011; Warren et Seifert, 2011; Merow *et al.*, 2013; Shcheglovitova et Anderson, 2013; Muscarella *et al.*, 2014; Cobos *et al.*, 2019c; Vignali *et al.*, 2020; Kass *et al.*, 2021). De plus, une trop grande complexité du modèle entraîne parfois une trop grande adéquation entre le modèle et les

<sup>5</sup> <https://github.com/mrmaxent/Maxent>.

données (*overfitting*). Ce comportement du modèle peut en effet limiter sa capacité à généraliser (e.g., Peterson *et al.*, 2007; Elith *et al.*, 2010), car celui-ci interprétera du bruit ou des biais comme faisant partie du motif recherché au lieu de ne pas les prendre en compte dans la formule (Peterson *et al.*, 2011; Merow *et al.*, 2014). Les causes de l'*overfitting* sont par exemple une trop grande aire de calibration par rapport à la répartition géographique des occurrences, mais également un paramétrage excessif (trop grande complexité du modèle) ou un trop faible nombre de d'occurrences.

Outre l'utilisation des *features*, Maxent va imposer des contraintes au modèle lui permettant de tenir compte des données d'occurrences tout en évitant l'*overfitting*. En effet, le modèle doit correspondre aux données mais également conserver sa capacité de généraliser, notamment dans l'objectif de transférer le modèle à une autre période ou une autre zone géographique. Dans le cas d'un *overfit* du modèle aux données, les courbes de réponse auront tendance à trop s'ajuster aux données de calibration, diminuant alors leur pouvoir prédictif. Il est donc nécessaire, d'une part, de lisser la distribution afin de limiter l'*overfitting* et, d'autre part, d'éliminer un certain nombre de features qui complexifieraient le modèle à outrance. Cette étape est contrôlée par un multiplicateur de régularisation L1 (parfois appelé « multiplicateur  $\beta$  »), dont le choix nécessite lui aussi une étape d'optimisation, la valeur par défaut n'étant pas toujours adéquate (e.g., Elith *et al.*, 2011; Warren *et al.*, 2011; Warren *et al.*, 2014; Merow *et al.*, 2013; Shcheglovitova *et al.*, 2013; Muscarella *et al.*, 2014; Radosavljevic *et al.*, 2014; Cobos *et al.*, 2019c; Vignali *et al.*, 2020; Kass *et al.*, 2021).

Pour terminer, la niche modélisée par l'algorithme Maxent et sa projection géographique se situera quelque part entre  $N_R$  et  $N_F^*$  en fonction de la configuration du diagramme **BAM** dans laquelle on se situe (Figure 4). Le premier type de niche sera approché dans la configuration de superposition: le modèle de sortie pourra être interprété comme une carte de probabilité d'occurrence (Peterson *et al.*, 2011, p. 109), puisque dans ce cas  $G_p$  équivaut à  $G_o$ . Dans les configurations classiques et de Hutchinson, il est possible, mais plus délicat d'estimer  $N_R$ , et donc  $G_o$ , car cela nécessite d'autres données (e.g., données de dispersion, vraies données d'absence; Peterson *et al.*, 2012). Il est donc plus probable que le modèle de sortie soit proche de  $N_F^*$ , étant donné qu'il

représenterait plutôt la probabilité de présence de conditions adéquates, alias favorabilité.

*Évaluation.* L'évaluation est une étape-clé dans la modélisation de niches par le biais d'algorithmes prédictifs tels que Maxent. C'est au cours de cette phase que l'on va déterminer si un modèle est robuste. Pour cela, il est nécessaire de quantifier deux aspects principaux: la performance et la signification.

La performance d'un modèle désigne sa capacité à atteindre un objectif précis dans l'absolu. Dans le cas de l'ENM, il s'agit de vérifier si le modèle prédit correctement des données d'occurrences indépendantes du jeu de données utilisé pour la calibration. Ce type de données est rarement disponible, et *a fortiori* impossible à obtenir pour des taxons paléontologiques ou des unités archéologiques. Plus généralement, on procède à un partitionnement du jeu de données d'occurrence initial afin de créer deux jeux de données supposés indépendants (Fielding *et al.*, 1997). Le but de cette manœuvre est tout d'abord de calibrer le modèle à partir de points de calibration (*model training* en anglais), puis de vérifier que les points de test sont correctement prédits par le modèle (*model testing*). Cette évaluation s'effectue par le biais de statistiques permettant de rendre compte de la capacité prédictive du modèle (i.e., sa capacité à prédire les points de test à partir des points de calibration), comme le taux d'erreur d'omission (Anderson *et al.*, 2003; Peterson *et al.*, 2008). Il est primordial que les points de test et de calibration soient suffisamment éloignés spatialement (donc potentiellement écologiquement) afin de ne pas artificiellement gonfler les mesures de performance. En d'autres termes, il ne faut pas que les points de test se situent dans les mêmes pixels que les points de calibration, afin de garantir l'indépendance des points de test et de calibration. Plusieurs méthodes de partitionnement existent et sont plus ou moins adaptées à des situations précises (e.g., Peterson *et al.*, 2011; Shcheglovitova *et al.*, 2013; Muscarella *et al.*, 2014; Radosavljevic *et al.*, 2014; Roberts *et al.*, 2017; Valavi *et al.*, 2018). Par exemple, la méthode du *jackknife* ou *leave-one-out* (Pearson *et al.*, 2006; Shcheglovitova *et al.*, 2013), qui consiste à évaluer la performance de chaque modèle calibré à partir des  $(n - 1)$  points d'occurrence avec le  $n^{\text{ième}}$  point, est adaptée pour les très petits jeux de données. Une autre méthode actuellement préconisée est le partitionnement par blocs spatiaux de validation croisée (Roberts *et al.*,

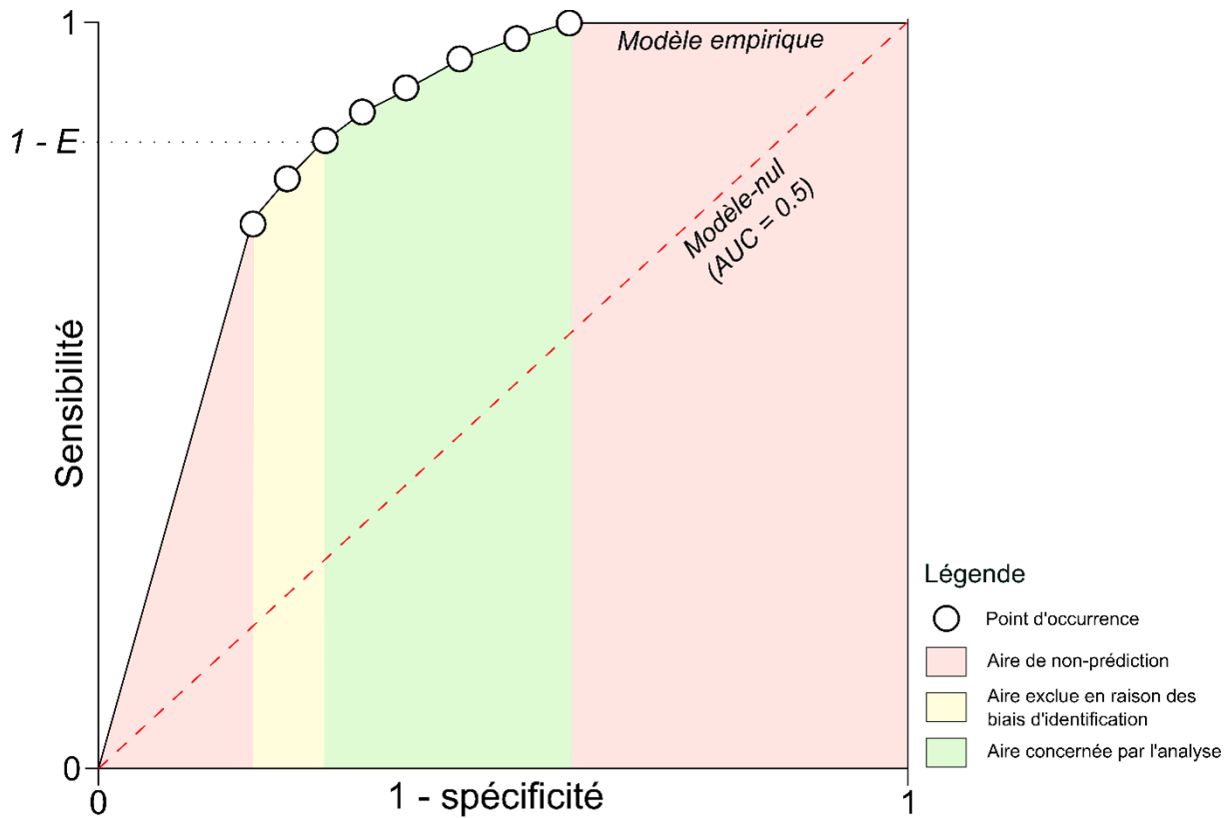


Figure 8. Exemple de *Receiver Operating Characteristic* (ROC) d'une prédiction Maxent pour évaluer sa signification statistique. La droite en pointillés rouges représente le ROC d'un modèle-nul non informatif et a donc une aire sous la courbe (AUC) de 0.5. La courbe noire représente le ROC du modèle empirique. Les zones colorées représentent son AUC. Les points blancs constituent les données servant à créer le modèle empirique. Dans l'approche du ROC partiel proposé par Peterson *et al.* (2008), les aires de non-prédiction (c'est-à-dire en dehors des données de calibration) en rouge ne sont pas prises en compte dans le calcul de l'AUC. D'autre part, l'aire sous la courbe correspondant à une proportion d'erreurs d'identification dans les données de calibration ( $E$ ) en jaune est également écartée du calcul (Figure modifiée d'après Peterson *et al.*, 2011, fig. 9.4, p.175).

2017; Valavi *et al.*, 2018), et permet notamment de prévenir l'autocorrélation spatiale entre les données de calibration et de test, garantissant ainsi mieux leur indépendance (Sillero et Barbosa 2020).

Déterminer la signification statistique d'un modèle consiste à évaluer si sa performance est meilleure que ce à quoi l'on pourrait s'attendre dans le cadre de l'hypothèse nulle. En d'autres termes, il s'agit de tester statistiquement si les prédictions des points d'évaluation par le modèle ne sont pas aléatoires compte tenu des données de calibration et d'évaluation (Peterson *et al.*, 2011). De la même façon que pour la mesure de la performance, il existe plusieurs méthodes pour tester la signification statistique d'un modèle (e.g., *ibid.*, Fielding et Bell, 1997; Peterson *et al.*, 2008), la plus couramment utilisée de nos jours étant le calcul de l'aire sous la courbe (*Area under curve*, abrégé. AUC) du *Receiver Operating Characteristic* (abrégé. ROC; e.g., Fielding et Bell, 1997). Le ROC est une courbe illustrant la variation de la sensibilité

du modèle (la proportion de présences connues correctement prédites, *i.e.*,  $1 -$  taux de faux négatifs) en fonction de  $1 -$  sa spécificité (la proportion d'absences connues prédites comme présentes, *i.e.*, taux de faux positifs). L'AUC du ROC est ensuite comparée à la courbe appartenant à un modèle non-informatif – dont l'AUC est égal à 0.5 – c'est-à-dire qu'il ne peut discriminer les vraies des fausses présences (Elith *et al.*, 2006; Peterson *et al.*, 2008; Figure 8). Or, lorsque le jeu de données n'est constitué que de présences, il n'est pas possible de correctement mesurer  $1 -$  la spécificité du modèle. Ainsi, il est préférable d'utiliser l'approche du ROC modifiée par Peterson *et al.* (2008), qui 1) s'adapte à l'approche présence/*background*, 2) se restreint au domaine prédit par l'algorithme sans se soucier des zones sujettes à l'extrapolation (en dehors de  $\mathbf{M}$ ), et 3) se restreint au domaine dans lequel le taux d'erreurs d'omissions ( $E$ ) est suffisamment faible en fonction d'une valeur prédéfinie par l'utilisateur.

Enfin, il faut également prêter une attention



particulière à la possibilité d'*overfitting* du modèle, c'est-à-dire une trop grande adéquation entre le modèle et les données. Celui-ci peut être en partie dû à une trop grande complexité du modèle liée au paramétrage. Dans le cadre de l'ENM, il est généralement préférable d'opter pour des modèles plus simples, car susceptibles de mieux refléter la réponse d'une espèce aux variables environnementales (Merow *et al.*, 2014, Jiménez *et al.*, 2019). De même que pour la performance et la signification, plusieurs méthodes existent pour sélectionner les modèles en fonction de leur complexité (*e.g.*, *ibid.*, Warren et Seifert, 2011; Warren *et al.*, 2014; Muscarella *et al.*, 2014; Cobos *et al.*, 2019c). La métrique la plus courante est le critère d'information d'Akaike (*Akaike information criteria*) corrigé en fonction de la taille de l'échantillon (abrév. AIC; Akaike, 1974; Warren et Seifert, 2011; Warren *et al.*, 2014). Cette métrique consiste en la standardisation des scores d'adéquation de telle sorte que la somme de ces scores dans l'espace géographique soit égale à 1. Ensuite, la probabilité des données par rapport au modèle est calculée par le produit des scores d'adéquation des pixels contenant une présence (*ibid.*).

*Choix du modèle final.* Une fois les meilleurs modèles déterminés par le protocole d'évaluation, deux approches principales permettent à l'utilisateur d'aboutir à un modèle final qui sera employé dans la suite des analyses (transferts et comparaisons). Certains défendent le choix du modèle obtenant les meilleures performances (*e.g.*, Qiao *et al.*, 2015). Cependant, cette approche comporte des biais, notamment liés à la difficulté de choisir un modèle parmi ceux ayant des performances égales ou similaires (Antunes, 2015). Il est donc souvent préconisé de créer un modèle de consensus de tous les modèles sélectionnés (*e.g.*, *ibid.*; Cobos *et al.*, 2019b, c). Cette méthode permet d'accéder à une estimation de l'incertitude associée à ces différents paramétrages d'un algorithme ou de différents algorithmes. Identifier les sources d'incertitudes et leur localisation dans l'espace géographique est une précaution importante pour l'interprétation des modèles de niches et pour les analyses ultérieures (transferts et comparaisons; Peterson *et al.*, 2018). Plusieurs méthodes de consensus existent (Antunes, 2015).

Enfin, le modèle final nécessite un dernier traitement, appelé seuillage, avant d'être interprété et analysé. Comme nous l'avons vu, les données

d'occurrences peuvent comporter des erreurs d'identification qu'il est nécessaire de prendre en compte dans la calibration du modèle. Cette prise en compte se fait notamment au moment d'évaluer la performance du modèle *via* le taux d'erreur d'omission. Par exemple, supposons que le taux d'erreurs d'identifications au sein d'un corpus soit d'environ 5 %. Il est alors nécessaire de permettre au modèle de se « tromper » (c'est-à-dire, d'omettre une occurrence dans la prédiction) dans maximum 5 % des cas (Peterson *et al.*, 2008). La mise en place de ce seuil doit nécessairement se retrouver dans le modèle final, par le biais du seuillage de la prédiction (Antunes, 2015; Sillero *et al.*, 2021). Dans mon exemple, nous pouvons considérer que la prédiction se trompe dans au plus 5 % des cas, et donc que 5 % des occurrences doivent se trouver en dehors des aires de favorabilité. De façon pratique, il s'agit de classer les occurrences en fonction de leur score de favorabilité, puis de retenir la valeur de favorabilité de la  $n^{\text{ième}}$  occurrence correspondant à 5 % du corpus. Tous les pixels dont le score de favorabilité est égal ou inférieur à cette valeur seront classés comme nuls (non adéquats; Figure 9). Cette méthode est dénommée le seuil de sensibilité fixe (*Fixed sensitivity*; Peterson *et al.*, 2011, p. 119), mais il existe d'autres façons de définir un seuil (*ibid.*).

*Modéliser et visualiser la niche fondamentale existante dans l'espace environnemental.* Malgré la popularité de Maxent, la forme des niches qu'il modélise reste généralement trop complexe pour correspondre à une approximation de  $N_F^*$  ou de  $N_F$  (Merow *et al.*, 2014). En effet, si la tolérance d'une espèce à une variable est assez vraisemblablement unimodale (Drake, 2015), alors  $N_F$  devrait théoriquement prendre une forme convexe, comme un ellipsoïde ou un polyèdre (*e.g.*, Van Aelst et Rousseeuw, 2009; Escobar *et al.*, 2014, 2017; Qiao *et al.*, 2016; Jiménez *et al.*, 2019; Soberón et Peterson, 2020; Nuñez-Penichet *et al.*, 2021b; Jiménez et Soberón, 2022). Ces dernières années, plusieurs études se sont concentrées sur l'emploi de l'ellipsoïde en ENM (Qiao *et al.*, 2016; Jiménez *et al.*, 2019; Nuñez-Penichet *et al.*, 2021b; Banks *et al.*, 2021; Vignoles, 2021; Jiménez et Soberón, 2022). Il s'agit en effet d'un objet simple à modéliser et à quantifier, car il comporte deux paramètres: le centroïde et la matrice de covariance (Jiménez *et al.*, 2019). La simplicité du modèle permet à la fois de réduire les hypothèses quant au paramétrage

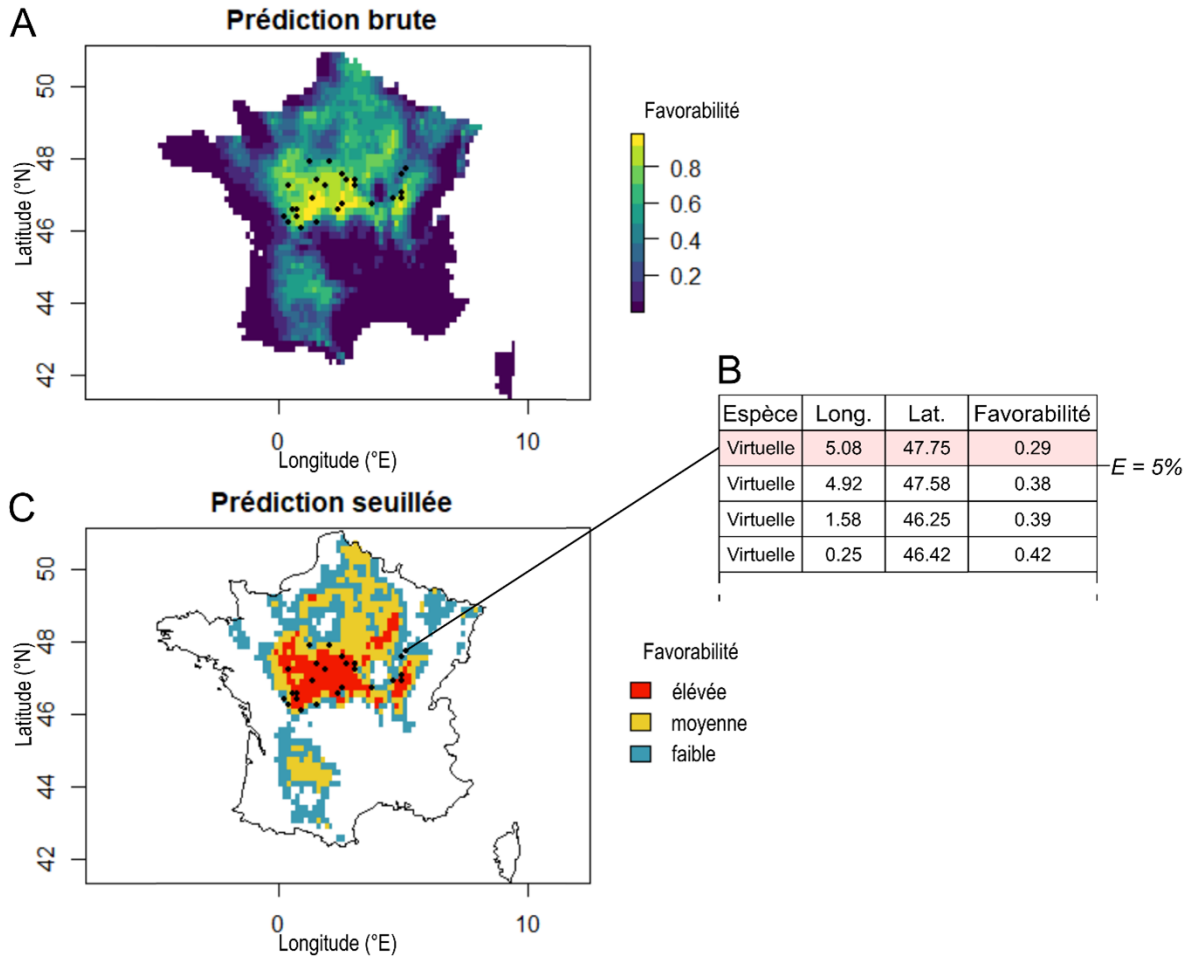


Figure 9. Illustration du seuillage de prédiction avec la méthode du seuil de sensibilité fixe. Le modèle de niche Maxent présenté correspond à celui d’une espèce virtuelle pour laquelle j’ai créé des données d’occurrences aléatoires (au nombre de 25). Les variables environnementales utilisées pour représenter l’espace environnemental sont la température moyenne annuelle, la précipitation moyenne annuelle et la saisonnalité de la précipitation modélisées pour la France de nos jours (1979-2013; modèle CHELSA, résolution spatiale de 10 arc min.; Karger *et al.*, 2017). J’ai employé le *package* R ENMTools (Warren *et al.*, 2021b) pour créer le modèle Maxent en utilisant les paramètres par défaut et en employant 20 % des données d’occurrences pour l’évaluation. Dans cet exemple, le seuil de sensibilité est fixé à  $E = 5\%$ . **A.** Prédiction « brute » à l’issue de la calibration: la favorabilité est présentée de manière continue. **B.** Quatre premières occurrences du jeu de données classées par favorabilité croissante. Les 5 % des occurrences associées aux plus faibles valeurs de favorabilité sont au nombre d’une seule; sa valeur de favorabilité est de 0.29 et sera retenue comme seuil pour le reclassement de la prédiction. **C.** Prédiction seuillée: les pixels aux valeurs inférieures ou égales à 0.29 sont considérés comme non-adéquats (blancs) tandis que les valeurs supérieures sont divisées en faible/moyenne/haute favorabilité.

(contrairement à Maxent; Merow *et al.*, 2013) et d’effectuer facilement des analyses comparatives du volume ou de la position des niches dans l’espace environnemental (Nuñez-Penichet *et al.*, 2021b, p. 4). En cela, ce type de modèle est plus adapté à des problématiques centrées sur les dynamiques de niches écologiques dans l’espace environnemental que des modèles plus complexes, comme Maxent – ces derniers seraient théoriquement plus pertinents pour modéliser  $N_R$  par exemple (Merow *et al.*, 2014).

#### Importance d’estimer l’aire accessible $M$ dans le processus d’ENM

La définition de  $M$  – la portion de paysage accessible à l’espèce étudiée en référence au diagramme **BAM** (Figure 3) – représente une étape majeure en ENM. C’est généralement l’aire qui est utilisée pour la calibration du modèle (Anderson et Raza, 2010; Barve *et al.*, 2011; Jiménez et Soberón, 2022). Elle va en effet représenter une zone dans laquelle l’espèce a pu se déplacer, donc dans laquelle son absence est *a priori* significative pour discriminer

les environnements favorables des environnements non-favorables (Barve *et al.*, 2011; Peterson *et al.*, 2011, p. 126). Toutefois, sa taille impacte grandement les résultats de la calibration et de l'évaluation (Lobo *et al.*, 2008; VanDerWal *et al.*, 2009; Anderson et Raza, 2010). En effet, si l'aire de calibration est trop petite et inférieure à **M**, le modèle peut sous-estimer les facteurs macro-géographiques qui influencent la répartition géographique de l'espèce – notamment **A** (Barve *et al.*, 2011). Lorsqu'elle est trop étendue, les statistiques d'évaluation sont exagérément bonnes, parce que l'étendue de la prédiction sera beaucoup plus importante. Pour le formuler autrement, il y a plus de chances que les points servant à tester la capacité prédictive du modèle soient prédits correctement, sans pour autant que la réponse des variables soit biologiquement cohérente (VanDerWal *et al.*, 2009). La définition de **M** influence aussi les étapes visant à transférer et comparer les modèles de niches (*e.g.*, Warren *et al.*, 2008; Owens *et al.*, 2013; Nuñez-Penichet *et al.*, 2021a, b), car ces deux opérations nécessitent d'être relativisées par rapport aux conditions environnementales présentes dans l'aire de calibration (donc **M**).

Pour guider le choix d'une aire de calibration qui représente au mieux les zones accessibles au taxon étudié (*i.e.*, **M**), il est nécessaire de se reposer sur des hypothèses biogéographiques claires plutôt que sur des limites administratives ou une quelconque autre aire arbitraire (Peterson *et al.*, 2011; Barve *et al.*, 2011). Ces hypothèses doivent prendre en compte la capacité dispersive de l'espèce considérée et notamment l'effet limitant de barrières naturelles (océan, montagne, fleuve...) empêchant l'accès des individus à des conditions favorables (*e.g.*, Barve *et al.*, 2011). Cette aire est généralement estimée manuellement, par exemple en traçant une zone de dispersion maximale autour des occurrences en fonction de ce qui est connu de l'espèce ou de la population étudiée et en écartant les régions séparées des occurrences par une barrière géographique (*e.g.*, Escobar *et al.*, 2018; Nuñez-Penichet *et al.*, 2021b; Vignoles, 2021). Une autre proposition est d'utiliser des régions biotiques ou biogéographiques – c'est-à-dire des régions présentant un cortège spécifique particulier et différent de régions voisines (Barve *et al.*, 2011), ou encore la distribution du ou des écotopes dans lesquels l'espèce a été observée (Soberón, 2010). Ces estimations restent subjectives, car elles ne permettent pas de précisément prendre

en compte l'histoire et les capacités dispersives du taxon. En effet, la période pendant laquelle l'espèce a été présente dans la zone d'étude impacte fortement ses possibilités d'accéder à des conditions environnementales favorables plus ou moins éloignées du cœur de sa répartition géographique (Barve *et al.*, 2011; Machado-Stredel *et al.*, 2021). Plus récemment, une méthode permettant d'estimer une aire **M** plus réaliste a été développée, en se basant sur des simulations de dispersion (*ibid.*). Cette approche permet de mieux approximer l'aire occupée (**G<sub>o</sub>**) pendant la durée d'existence de l'espèce dans une région, en fonction des fluctuations climatiques qu'elle a traversées et qui ont certainement influencé son accessibilité à certaines zones.

Dans le cas de l'application de l'ENM au registre fossile ou archéologique, la définition de l'aire de calibration **M** peut être un défi d'autant plus grand. En effet, les connaissances limitées des comportements de dispersion d'organismes ou de populations aujourd'hui disparues obligent à se reposer sur des hypothèses d'autant plus spéculatives. Certaines propositions existent toutefois dans la littérature. Par exemple, dans le cas d'espèces fossiles, il est suggéré d'employer comme aire de calibration l'étendue spatiale des couches affleurantes dans laquelle l'organisme a été échantillonné (Myers *et al.*, 2015). En dehors de ces zones, il n'est en effet pas possible de savoir si les conditions environnementales étaient ou non favorables à l'implantation de l'organisme. Concernant l'application de l'ENM aux données archéologiques, l'estimation de **M** reste un point méthodologique encore peu exploré. Les études récentes se basent sur des hypothèses biogéographiques – par exemple, en excluant les zones recouvertes de glaciers ou en incluant des portions de continent alors émergés – mais aussi culturelles, en se référant notamment au principe d'actualisme (*i.e.*, les distances maximales parcourues par des groupes de chasseurs-cueilleurs subactuels; Vignoles *et al.*, 2021) ou à la provenance de matériaux présents au sein des sites (silex, parures en coquillages; Vignoles, 2021). Ces estimations de **M** restent toutefois approximatives. Il conviendrait à l'avenir d'améliorer cette étape par le développement de méthodologies conduisant à des hypothèses de dispersion plus réalistes, en s'inspirant par exemple de travaux qui cherchent à modéliser les territoires parcourus par des groupes de chasseurs-cueilleurs par le biais de la modélisation d'agents (*agent-*

*based modeling*) et/ou du chemin le moins coûteux (*least cost path*; e.g., Gravel-Miguel et Wren, 2018; Vaissié, 2021).

*Transfert du modèle à d'autres régions ou périodes: précautions*

Une fois le modèle calibré dans **M**, il est souvent nécessaire de projeter ce modèle dans d'autres contrées ou d'autres périodes. Cette opération peut se révéler assez délicate en raison de la présence d'environnements inconnus à l'aire de calibration dans la nouvelle zone, ou dans la même zone mais sous un climat différent (e.g., Randin *et al.*, 2006; Williams et Jackson, 2007; Williams *et al.*, 2007; Zurell *et al.*, 2012; Owens *et al.*, 2013; Peterson *et al.*, 2018). Ces conditions non-analogues peuvent être dues à la présence de nouvelles valeurs – par exemple, les valeurs de température moyenne annuelle varient entre 5°C et 10°C dans l'aire de calibration, tandis qu'elles varient entre 10°C et 20°C dans la zone où le modèle sera projeté (Williams *et al.*, 2007) – ou à la présence d'une nouvelle combinaison entre variables – par exemple, la combinaison de 5°C de température moyenne annuelle avec 350 mm.an<sup>-1</sup> de précipitation moyenne annuelle existe dans la zone de projection mais pas dans la zone de calibration (Zurell *et al.*, 2012). Or, ces conditions n'ayant pas été prises en compte dans la calibration du modèle, ce dernier sera obligé d'extrapoler lorsqu'il y sera confronté (Owens *et al.*, 2013; Peterson *et al.*, 2011). Les prédictions du modèle dans ces conditions nouvelles auront de grandes chances d'être hautement incertaines, car le comportement du modèle ne peut être vérifié ou anticipé au-delà des conditions de calibration (Figure 10). La réponse d'un individu à la variation d'une variable environnementale peut être représentée comme une courbe unimodale. Or, lorsque cette courbe est tronquée en raison de l'absence d'une partie des valeurs de la variable environnementale dans l'aire de calibration, il est impossible de déterminer si la valeur sélective (*fitness*) va augmenter, rester constante ou diminuer (Figure 10). La troncature des courbes de réponse est assez fréquente (Peterson *et al.*, 2011, 2018) puisque la modélisation corrélatrice de niches ne permet d'accéder qu'à des portions de  $N_F$  ( $N_F^*$  ou  $N_R$  en fonction des approches et données). Malheureusement, les approches corrélatrices ne sont pas forcément idéales pour déterminer des relations mécanistiques entre occurrences et environnement,

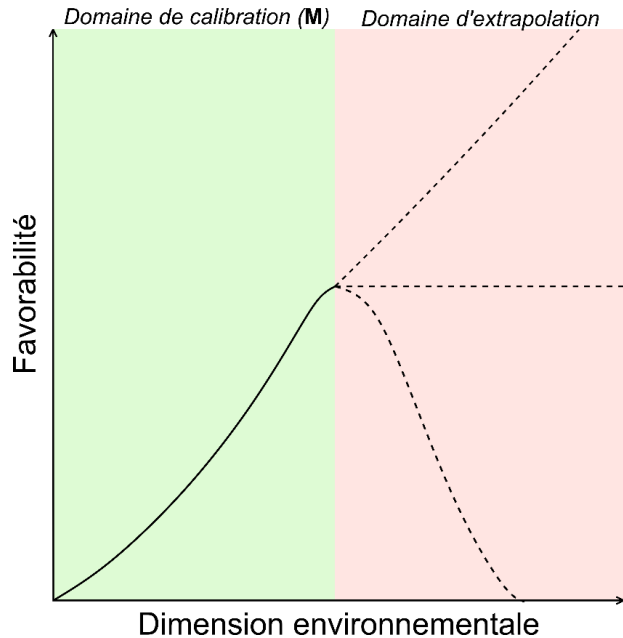


Figure 10. Illustration du risque d'extrapolation de prédictions de niches hors de la gamme environnementale des données de calibration (domaine vert). En dehors de la gamme environnementale de calibration (*i.e.*, non représentée dans la région **M**, domaine rouge), il n'est pas possible de vérifier si la courbe de réponse (en pointillés) va continuer à monter, rester constante ou redescendre. Il s'agit donc d'une courbe de réponse tronquée (Figure et légende modifiées d'après Peterson *et al.*, 2011, fig.7.6, p. 127).

ce qui rend l'extrapolation en dehors de **M** incertaine (Peterson *et al.*, 2011; Owens *et al.*, 2013).

Afin d'opérer un transfert limitant l'extrapolation du modèle, il est important d'évaluer la transférabilité du modèle à une autre zone ou une autre période (Elith *et al.*, 2010; Zurell *et al.*, 2012; Owens *et al.*, 2013). Il s'agit de cartographier les conditions analogues (*i.e.*, similaires) à ce qui existe dans **M** dans ce nouvel environnement. La métrique la plus récemment développée s'appelle *Mobility-Oriented Parity* (abrév. MOP) et « *identifie les aires de stricte extrapolation et calcule la similarité environnementale entre les régions de calibration et de projection* » (Owens *et al.*, 2013, p. 13). Cette méthode se base sur le calcul de distances multivariées (e.g., distance de Mahalanobis; Mahalanobis, 1936; Etherington, 2019) entre les environnements associés aux points de la région de projection et une proportion (prédéfinie par l'utilisateur) des environnements associés aux points de la région de calibration. Cette proportion est généralement restreinte à une portion du nuage de points de **M** proche du nuage de points correspondant à la région de projection (*ibid.*). La carte de répartition des environnements similaires



et des zones d'extrapolation stricte issue du MOP pourra ensuite être superposée aux prédictions de niches, afin d'identifier les valeurs de favorabilité relevant de l'extrapolation pure et simple.

#### *Comment comparer deux modèles de niches ?*

La comparaison de modèles de niches est une analyse très couramment mise en œuvre en ENM, car elle permet d'aborder de nombreuses problématiques ayant trait à l'évolution des espèces ou à leurs similarités écologiques. Il s'agit toutefois d'une étape délicate, car le choix d'une méthode de comparaison peut largement influencer les résultats (e.g., Warren *et al.*, 2008; Rödder et Engler, 2011). De façon générale, il est important de prendre en compte la dualité de Hutchinson lors de cette étape, c'est-à-dire de proposer des comparaisons à la fois dans l'espace géographique et environnemental, étant donné que les propriétés des niches dans ces deux espaces peuvent être très différentes (Figure 2).

Dans l'espace environnemental, il existe plusieurs approches en fonction de si l'on cherche à comparer leur recouvrement (*i.e.*, les niches partagent-elles une partie de leurs environnements ?), leur proximité (*i.e.*, la distance entre les niches dans l'espace multivarié des niches est-elle petite ?; Mammola, 2019), ou encore leur similarité (*i.e.*, les niches présentent-elles une forme similaire ?), leur équivalence (*i.e.*, les niches sont-elles parfaitement identiques d'un point de vue environnemental?; Warren *et al.*, 2008). Les deux premiers types de comparaisons se focalisent sur la localisation relative des niches au sein de l'espace environnemental et donnent généralement des résultats similaires (sauf dans le cas de volumes totalement disjoints: les mesures de recouvrement sont alors inefficaces; Mammola, 2019). En revanche, les deux autres types de comparaison se focalisent sur la forme des enveloppes, sans référence à leur localisation dans l'espace environnemental, ce qui permet alors de répondre à d'autres types de questions.

Dans l'espace géographique, plusieurs types de comparaisons sont également proposés pour analyser la similarité de la distribution géographique de modèles de niches (e.g., Warren *et al.*, 2008; Broennimann *et al.*, 2011; Nuñez-Penichet *et al.*, 2021a). Ces méthodes vont se concentrer sur la comparaison de distributions de probabilités ou de favorabilité données par un ou des algorithmes; il ne s'agit donc pas de comparer des niches d'un

point de vue environnemental, mais plutôt d'évaluer la similarité de la répartition d'environnements favorables ou occupés par les deux espèces considérées. La méthodologie la plus simple consiste à binariser les prédictions entre les pixels favorables et non-favorables, puis à la soustraire afin d'identifier les zones présentant un gain, une perte ou une absence de changement de la favorabilité (e.g., Nuñez-Penichet *et al.*, 2021a; Banks *et al.*, 2021). Elle ne permet toutefois pas de comparer le degré de favorabilité de différentes zones. Pour ce faire, des statistiques plus complexes doivent être utilisées, qui prennent en compte les scores de favorabilité dans la comparaison (e.g., Warren *et al.*, 2008; Broennimann *et al.*, 2011).

Quelque soit le type de comparaison effectué, la signification des valeurs obtenues doit ensuite être vérifiée statistiquement (Figure 11). En particulier, la proportion des environnements partagés par l'aire de calibration des deux modèles joue un rôle crucial dans la pertinence de la comparaison (e.g., Warren *et al.*, 2008; Banks *et al.*, 2021). Si les environnements accessibles à deux espèces comparées sont trop différents, l'absence ou quasi-absence de recouvrement/similarité entre leurs niches sera peu, voire pas significative, car la comparaison de niches modélisées à partir des données d'occurrences prélevées aléatoirement dans les aires de calibration respectives aurait donné le même résultat (par exemple, Figure 11B). En d'autres termes, l'absence de recouvrement résulterait de la chance plutôt que d'une véritable différence entre les niches par rapport aux environnements disponibles aux deux espèces dans leurs **M** respectifs. À l'inverse, si les environnements accessibles présentent un recouvrement important, mais que les modèles de niches ne se recouvrent que peu, alors celle-ci sera probablement significative étant donné que des niches modélisées à partir de données prélevées aléatoirement dans les aires de calibration respectives ont plus de chances de se recouvrir. La signification d'une comparaison est généralement testée par le biais de tests de randomisation (Figure 11; e.g., Warren *et al.*, 2008, 2021; Nuñez-Penichet *et al.*, 2021b; Banks *et al.*, 2021). Ceux-ci consistent à comparer la valeur empirique de la métrique employée avec une distribution nulle de valeurs nulles mesurées pour la comparaison de paires de modèles aléatoires générés à partir du *background* (*i.e.*, environnements accessibles à chaque espèce comparée). Afin de

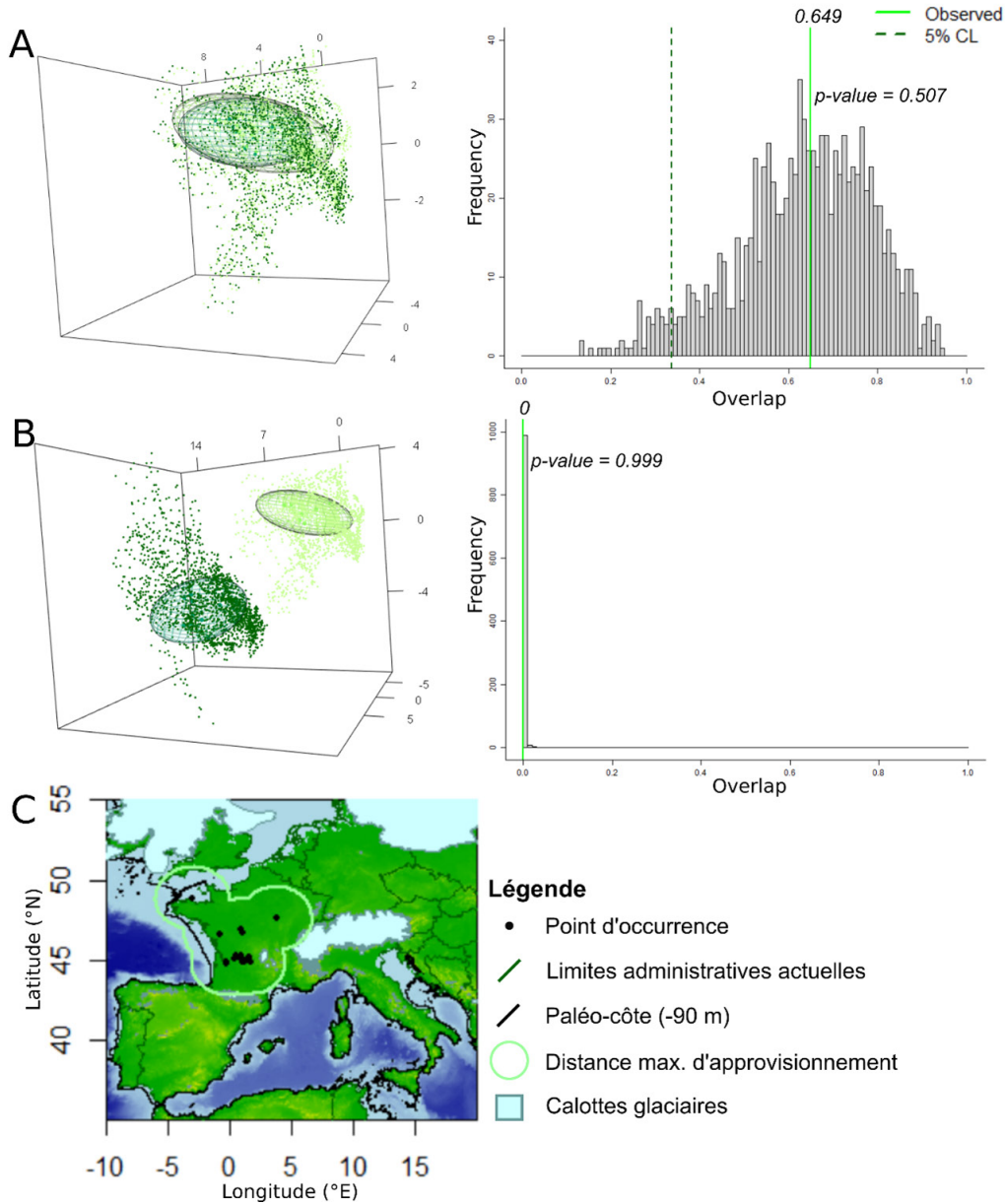


Figure 11. Exemples de comparaisons du recouvrement d'ellipsoïdes en employant l'indice de Jaccard (*cf.* Mammola, 2019), dans le cadre de l'évaluation de l'impact du choix d'un type de simulation paléoclimatique sur les modèles de niches associés à une culture archéologique du Paléolithique supérieur appelée Rayssien (31 900 – 26 900 ans calibrés avec le présent; Vignoles, 2021). Cette culture est caractérisée par une méthode de fabrication d'armes de chasse particulière, qui n'était employée que dans une aire géographique restreinte et des environnements relativement spécifiques. Ces analyses ont été réalisées avec le *package* R *ellipsenm*<sup>5</sup> (*e.g.*, Nuñez-Penichet *et al.*, 2021b; Banks *et al.*, 2021). **A.** Comparaison d'ellipsoïdes modélisés à partir de deux jeux de simulations paléoclimatiques transitoires (Armstrong *et al.*, 2019) moyennées sur 30 ans (vert clair) et 100 ans (vert foncé). Le recouvrement des ellipsoïdes est important mais n'est pas significatif au regard de la proportion d'environnements partagés (fort recouvrement entre les *background*). **B.** Comparaison d'ellipsoïdes modélisés à partir d'un jeu de simulations paléoclimatiques transitoires (vert clair; Armstrong *et al.*, 2019) et d'un jeu de simulations paléoclimatiques calculées à l'équilibre (vert foncé; Beyer *et al.*, 2020). Le recouvrement des ellipsoïdes est nul, mais n'est pas significatif au regard de la proportion d'environnements partagés (quasi-absence de recouvrement entre les *background*). **C.** Aire de calibration définie par l'intersection entre le trait de paléo-côte (-90 m il y a *ca.* 30 000 ans; Siddall *et al.*, 2003), l'extension maximale des calottes glaciaires (Ehlers et Gibbard, 2004) et la distance maximale d'approvisionnement en silex identifiée pour le Rayssien (220 km; Vignoles, 2021); points d'occurrences du Rayssien.

<sup>5</sup> <https://github.com/marloncobos/ellipsenm/>.

rejeter l'hypothèse de similarité (ou de différence), il faut alors que la valeur empirique se situe en dessous du seuil de signification défini par l'utilisateur (généralement 5%, *i.e.*,  $p < 0.05$ )

#### CONCLUSIONS

Dans cet article, j'ai présenté un cadre théorique dans lequel inscrire toute étude en ENM. J'ai pour cela rappelé les principaux concepts et terminologie de la théorie des niches, ainsi que sa relation avec la distribution géographique d'un taxon ou une population. Les différentes étapes pratiques de l'ENM ont ensuite été décrites: la construction de jeux de données d'occurrences et le choix des données environnementales; la création et l'évaluation des modèles de niches; leur transfert à d'autres régions ou périodes et leur comparaison.

Bien que loin d'être exhaustive, cette synthèse récapitule les principaux points de vigilance et éléments à prendre en compte lors de la mise en œuvre de l'ENM, quel que soit le contexte d'application. J'ai particulièrement insisté sur (1) l'importance d'un regard critique sur les données de départ. Les données d'occurrences peuvent être affectées par de nombreux biais qu'il convient d'identifier et de prendre en compte. De même, les données environnementales sont majoritairement issues de modèles et comportent un certain degré d'incertitude. Ma synthèse a également mis en évidence (2) la diversité de pratiques qui existent à toutes les étapes de la chaîne. Que ce soit dans le choix des prédicteurs environnementaux, de l'algorithme, de l'aire de calibration, de l'évaluation ou du paramétrage du modèle, de nombreuses approches peuvent être adoptées. C'est pourquoi je préconise, dans la mesure du possible, d'évaluer et de présenter la variabilité découlant de tout ou partie de ces choix méthodologiques sur les modèles finaux et leurs interprétations (Diniz-Filho *et al.*, 2009; Warren *et al.*, 2021a). Enfin, une fois les modèles calibrés, (3) leur transfert et leurs comparaisons nécessitent des précautions afin d'éviter que ces analyses et leurs interprétations soient trop fortement biaisées. Ces opérations reposent en effet sur une référence à l'aire de calibration du modèle, qui doit servir à relativiser les comparaisons et tester leur significativité.

Ces préconisations se fondent sur un principe à garder à l'esprit lors de l'emploi de modèles mathématique: leurs résultats ne sont qu'une approximation du réel. Ils dépendent majoritairement

des données et des méthodes mathématiques de création et d'analyse des modèles. C'est pourquoi une vigilance constante quant à l'adéquation des données et des choix méthodologiques avec le cadre théorique de référence et les problématiques de l'étude est primordiale pour modéliser au mieux le phénomène étudié.

#### REMERCIEMENTS

Je remercie A. Townsend Peterson de m'avoir proposé d'écrire cet article pour *Biodiversity Informatics*. Merci également à Amélie Challier pour sa relecture de la première version du manuscrit et ses suggestions d'amélioration. Merci enfin à Lucas Buffan et un relecteur anonyme pour leurs commentaires et suggestions qui m'ont permis de clarifier et de compléter cette synthèse. L'article est une adaptation de plusieurs chapitres de ma thèse de doctorat, dont l'élaboration a été financée par le projet Région Nouvelle-Aquitaine « GravettoNiches » (dir. W. E. Banks)

#### DATA AVAILABILITY

Les scripts R et les données ayant permis la réalisation de figures-exemples 2 et 9 sont accessibles à / The R scripts and data used for example-figures 2 and 9 are accessible at [https://osf.io/3nwqk/?view\\_only=7e51054abaf74582a7701253ec37704f](https://osf.io/3nwqk/?view_only=7e51054abaf74582a7701253ec37704f)

#### DECLARATION OF COMPETING INTERESTS

The author has declared that no competing interests exist.

#### LITERATURE CITED

- Aiello-Lammens, M. E., R. A. Boria, A. Radosavljevic, B. Vilela, et R. P. Anderson. 2015. SpThin: An R Package for Spatial Thinning of Species Occurrence Records for Use in Ecological Niche Models. *Ecography* 38 (5): 541-45.
- Akaike, H. 1974. A New Look at the Statistical Model Identification. *IEEE Trans. Automat. Contr.* AC-19 (6): 716-23.
- Alkische, A., M. E. Cobos, A. T. Peterson, et A. M. Samy. 2020. Recognizing Sources of Uncertainty in Disease Vector Ecological Niche Models: An Example with the Tick *Rhipicephalus sanguineus sensu lato*. *Perspect. Ecol. Conserv.* 18 (2): 91-102.
- Amatulli, G., S. Domisch, M. Tuanmu, B. Parmentier, A. Ranipeta, J. Malczyk, et W. Jetz. 2018. A Suite of Global, Cross-Scale Topographic Variables for Environmental and Biodiversity Modeling. *Sci. Data* 5 (1): 180040.
- Amatulli, G., D. McInerney, T. Sethi, P. Strobl, et S. Domisch. 2020. Geomorpho90m, Empirical Evaluation and Accuracy

- Assessment of Global High-Resolution Geomorphometric Layers. *Sci. Data* 7 (1): 162.
- Anderson, R. P., et I. Jr. Gonzalez. 2011. Species-Specific Tuning Increases Robustness to Sampling Bias in Models of Species Distributions: An Implementation with Maxent. *Ecol. Modell.*, 222: 279-2811.
- Anderson, R. P., et A. Raza. 2010. The Effect of the Extent of the Study Region on GIS Models of Species Geographic Distributions and Estimates of Niche Evolution: Preliminary Tests with Montane Rodents (Genus *Nephelomys*) in Venezuela. *J. of Biogeogr.* 37 (7): 1378-93.
- Anderson, R. P., D. Lew, et A. T. Peterson. 2003. Evaluating Predictive Models of Species' Distributions: Criteria for Selecting Optimal Models. *Ecol. Modell.* 162 (3): 211-32.
- Angilletta, M. J. 2009. *Thermal Adaptation: A Theoretical and Empirical Synthesis*. Oxford University Press, Oxford.
- Antunes, N. 2015. Application d'algorithmes prédictifs à l'identification de niche écoculturelles des populations du passé : approche ethnoarchéologique. Ph.D thesis, Université de Bordeaux, Pessac (France).
- Araújo, M. B., et A. Guisan. 2006. Five (or so) Challenges for Species Distribution Modelling. *J. of Biogeogr.* 33 (10): 1677-88.
- Araújo, M. B., et A. T. Peterson. 2012. Uses and Misuses of Bioclimatic Envelope Modeling. *Ecology* 93 (7): 1527-39.
- Araújo, M. B., et A. Rozenfeld. 2013. The Geographic Scaling of Biotic Interactions. *Ecography* 37 (5): 406-15.
- Araújo, M. B., R. P. Anderson, A. M. Barbosa, C. M. Beale, C. F. Dormann, R. Early, R. A. Garcia, et al. 2019. Standards for Distribution Models in Biodiversity Assessments. *Sci. Adv.* 5 (1): eaat4858.
- Ashraf, U., A. T. Peterson, M. N. Chaudhry, I. Ashraf, Z. Saqib, S. R. Ahmad, et H. Ali. 2017. Ecological Niche Model Comparison under Different Climate Scenarios: A Case Study of *Olea* spp. in Asia. *Ecosphere* 8 (5): e01825.
- Armstrong, E., P. O. Hopcroft, et P. J. Valdes. 2019. A Simulated Northern Hemisphere Terrestrial Climate Dataset for the Past 60,000 Years. *Sci. Data* 6 (1): 265.
- Banks, W. E., F. d'Errico, H. I. Dibble, L. Krishtalka, D. West, D. I. Olszewski, A. T. Peterson, et al. 2006. Eco-Cultural Niche Modeling: New Tools for Reconstructing the Geography and Ecology of Past Human Populations. *PaleoAnthropology*, 68–83.
- Banks, W. E., J. Zilhão, F. d'Errico, M. Kageyama, A. Sima, et A. Ronchitelli. 2009. Investigating Links between Ecology and Bifacial Tool Types in Western Europe during the Last Glacial Maximum. *J. Archaeol. Sci.* 36 (12): 2853-67.
- Banks, W. E., T. Aubry, F. d'Errico, J. Zilhão, A. Lira-Noriega, et A. T. Peterson. 2011. Eco-Cultural Niches of the Badegoulian: Unraveling Links between Cultural Adaptation and Ecology during the Last Glacial Maximum in France. *J. Anthropol. Archaeol.* 30 (3): 359-74.
- Banks, W. E., M.-H. Moncel, J.-P. Raynal, M. E. Cobos, D. Romero-Alvarez, M.-N. Woillez, J.-P. Faivre, et al. 2021. An Ecological Niche Shift for Neanderthal Populations in Western Europe 70,000 Years Ago. *Sci. Rep.* 11 (1): 5346.
- Barve, N., V. Barve, A. Jiménez-Valverde, A. Lira-Noriega, S. P. Maher, A. T. Peterson, J. Soberón, et F. Villalobos. 2011. The Crucial Role of the Accessible Area in Ecological Niche Modeling and Species Distribution Modeling. *Ecol. Modell.* 222 (11): 1810-19.
- Beyer, R. M., M. Krapp, et A. Manica. 2020. An Empirical Evaluation of Bias Correction Methods for Palaeoclimate Simulations. *Clim. Past* 16 (4): 1493-1508.
- Booth, T. H., H. A. Nix, J. R. Busby, et M. F. Hutchinson. 2014. BIOCLIM : The First Species Distribution Modelling Package, Its Early Applications and Relevance to Most Current MAXENT Studies. *Diversity Distrib.* 20 (1): 1-9.
- Boria, R. A., L. E. Olson, S. M. Goodman, et R. P. Anderson. 2014. Spatial Filtering to Reduce Sampling Bias Can Improve the Performance of Ecological Niche Models. *Ecol. Modell.* 275: 73-77.
- Boucher, O., J. Servonnat, A. L. Albright, O. Aumont, Y. Balkanski, V. Bastrikov, S. Bekki, et al. 2020. Presentation and Evaluation of the IPSL-CM6A-LR Climate Model. *J. Adv. Model. Earth Syst.* 12 (7): e2019MS002010.
- Breiman, L. 2001. Random Forests. *Mach. Learn.* 45: 5-32.
- Breiner, F. T., A. Guisan, A. Bergamini, et M. P. Nobis. 2015. Overcoming Limitations of Modelling Rare Species by Using Ensembles of Small Models. *Methods Ecol. Evol.* 6 (10): 1210-18.
- Broennimann, O., M. C. Fitzpatrick, P. B. Pearman, B. PetitP., L. Pellissier, N. G. Yoccoz, W. Thuiller, et al. 2011. Measuring ecological niche overlap from occurrence and spatial environment data. *Glob. Ecol. Biogeogr.* 21 (4): 481-97.
- Carpenter, G., A. N. Gillison, et J. Winter. 1993. DOMAIN: A Flexible Modelling Procedure for Mapping Potential Distributions of Plants and Animals. *Biodivers. Conserv.* 2 (6): 667-80.
- Chapman, A., J. Wiczorek, P. Zermoglio, M. Luna, et D. Bloom. 2020. Improved Georeferencing: Three Essential Guiding Documents. *Biodiv. Information Sci. Stand.* 4: e58983.
- Chase, J. M., et M. A. Leibold. 2003. *Ecological Niches: Linking Classical and Contemporary Approaches*. Interspecific Interactions. University of Chicago Press, Chicago.
- Claussen, M., L. Mysak, A. Weaver, M. Crucifix, T. Fichefet, M.-F. Loutre, S. Weber, et al. 2002. Earth System Models of Intermediate Complexity: Closing the Gap in the Spectrum of Climate System Models. *Clim. Dyn.* 18 (7): 579-86.
- Cobos, M. E., A. T. Peterson, L. Osorio-Olvera, et D. Jiménez-García. 2019a. An Exhaustive Analysis of Heuristic Methods for Variable Selection in Ecological Niche Modeling and Species Distribution Modeling. *Ecol. Inform.* 53: 100983.
- Cobos, M. E., L. Osorio-Olvera, et A. T. Peterson. 2019b. Assessment and Representation of Variability in Ecological Niche Model Predictions. *BioRxiv*, 603100.
- Cobos, M. E., A. T. Peterson, N. Barve, et L. Osorio-Olvera. 2019c. KUENM: An R Package for Detailed Development of Ecological Niche Models Using Maxent. *PeerJ* 7: e6281.
- Colwell, R. K., et T. F. Rangel. 2009. Hutchinson's Duality: The Once and Future Niche. *Proc. Natl. Acad. Sci. U.S.A.* 106 (2): 19651-58.



- Diniz-Filho, J. A. F., L. M. Bini, T. F. Rangel, R. D. Loyola, C. Hof, D. Nogués-Bravo, et M. B. Araújo. 2009. Partitioning and Mapping Uncertainties in Ensembles of Forecasts of Species Turnover under Climate Change. *Ecography* 32 (6): 897-906.
- Domisch, S., G. Amatulli, et W. Jetz. 2015. Near-Global Freshwater-Specific Environmental Variables for Biodiversity Analyses in 1 Km Resolution. *Sci. Data* 2 (1): 150073.
- Drake, J. M. 2015. Range Bagging: A New Method for Ecological Niche Modelling from Presence-Only Data. *J. R. Soc. Interface* 12 (107): 20150086.
- Ehlers, J., et P. Gibbard. 2004. Quaternary Glaciations - Extent and Chronology, Part I: Europe. Elsevier, Amsterdam.
- Elith, J., C. H. Graham, R. P. Anderson, M. Dudík, Simon Ferrier, A. Guisan, R. J. Hijmans, et al. 2006. Novel Methods Improve Prediction of Species' Distributions from Occurrence Data. *Ecography* 29 (2): 129-51.
- Elith, J., M. Kearney, et S. Phillips. 2010. The Art of Modelling Range-Shifting Species. *Methods Ecol. Evol.* 1 (4): 330-42.
- Elith, J., S. J. Phillips, Trevor Hastie, M. Dudík, Yung En Chee, et C. J. Yates. 2011. A Statistical Explanation of MaxEnt for Ecologists. *Divers. Distrib.* 17 (1): 43-57.
- Elton, C. 1927. *Animal Ecology*. The MacMillan Company, New York.
- Escobar, L. E., A. Lira-Noriega, G. Medina-Vogel, et A. T. Peterson. 2014. Potential for Spread of the White-Nose Fungus (*Pseudogymnoascus destructans*) in the Americas: Use of Maxent and NicheA to Assure Strict Model Transference. *Geospat. Health* 9 (1): 221-29.
- Escobar, L. E., H. Qiao, C. Lee, et N. B. D. Phelps. 2017. Novel Methods in Disease Biogeography: A Case Study with Heterosporosis. *Front. Vet. Sci.* 4: 105.
- Escobar, L. E., H. Qiao, J. Cabello, et A. T. Peterson. 2018. Ecological Niche Modeling Re-Examined: A Case Study with the Darwin's Fox. *Ecol. Evol.* 8 (10): 4757-70.
- Etherington, T. R. 2019. Mahalanobis Distances and Ecological Niche Modelling: Correcting a Chi-Squared Probability Error. *PeerJ* 7: e6678.
- Farber, O., et R. Kadmon. 2003. Assessment of Alternative Approaches for Bioclimatic Modeling with Special Emphasis on the Mahalanobis Distance. *Ecol. Modell.* 160 (1-2): 115-30.
- Feng, Xiao, D. S. Park, Cassondra Walker, A. T. Peterson, Cory Merow, et M. Papeş. 2019a. A Checklist for Maximizing Reproducibility of Ecological Niche Models. *Nat. Ecol. Evol.* 3 (10): 1382-95.
- Feng, X., D. S. Park, Y. Liang, R. Pandey, et M. Papeş. 2019b. Collinearity in Ecological Niche Modeling: Confusions and Challenges. *Ecol. Evol.* 9 (18): 10365-76.
- Fick, S. E., et R. J. Hijmans. 2017. WorldClim 2: New 1-km Spatial Resolution Climate Surfaces for Global Land Areas. *Int. J. Climatol.* 37 (12): 4302-15.
- Fielding, A. H., et J. F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* 24 (1): 38-49.
- Gause, G. F. 1934. *The Struggle for Existence*. The Williams & Wilkins Company, Baltimore.
- Gibert, C., A. Vignoles, C. Contoux, W. E. Banks, D. Barboni, J.-R. Boisserie, O. Chavasseau, et al. 2022. Climate-Inferred Distribution Estimates of Mid-to-Late Pliocene Hominins. *Glob. Planet. Change* 210: 103756.
- Goosse, H., V. Brovkin, T. Fichefet, R. Haarsma, P. Huybrechts, J. Jongma, A. Mouchet, et al. 2010. Description of the Earth System Model of Intermediate Complexity LOVECLIM Version 1.2. *Geosci. Model Dev.* 3 (2): 603-33.
- Gravel-Miguel, C., et C. D. Wren. 2018. Agent-Based Least-Cost Path Analysis and the Diffusion of Cantabrian Lower Magdalenian Engraved Scapulae. *J. Archaeol. Sci.* 99: 1-9.
- Green, R. H. 1971. A Multivariate Statistical Approach to the Hutchinsonian Niche: Bivalve Molluscs of Central Canada. *Ecology* 52 (4): 543-56.
- Grinnell, J. 1917. The Niche-Relationships of the California Thrasher. *Auk* 34 (4): 427-33.
- Grubb, P. J. 1977. The Maintenance of Species-Richness in Plant Communities: The Importance of the Regeneration Niche. *Biol. Rev.* 52 (1): 107-45.
- Guisan, A., T. C. Edwards, et T. Hastie. 2002. Generalized Linear and Generalized Additive Models in Studies of Species Distributions: Setting the Scene. *Ecol. Modell.* 157 (2-3): 89-100.
- Guisan, A., et N. E. Zimmermann. 2000. Predictive Habitat Distribution Models in Ecology. *Ecol. Modell.* 135 (2-3): 147-86.
- Hengl, T., J. Mendes de Jesus, G. B. M. Heuvelink, M. Ruiperez Gonzalez, M. Kilibarda, A. Blagotić, W. Shangquan, et al. 2017. SoilGrids250m: Global Gridded Soil Information Based on Machine Learning. *PLoS ONE* 12 (2): e0169748.
- Hernandez, P. A., C. H. Graham, L. L. Master, et D. L. Albert. 2006. The Effect of Sample Size and Species Characteristics on Performance of Different Species Distribution Modeling Methods. *Ecography* 29 (5): 773-85.
- Hutchinson, G. E. 1957. Population studies: Animal ecology and demography. *Bull. Math. Biol.* 53 (1-2): 193-213.
- Ingenloff, K. 2020. Enhancing the Correlative Ecological Niche Modeling Framework to Incorporate the Temporal Dimension of Species' Distributions. Ph.D thesis, University of Kansas, Lawrence.
- Ingenloff, K., et A. T. Peterson. 2021. Incorporating time into the traditional correlational distributional modeling framework: a proof-of-concept using the Wood Thrush (*Hylocichla mustelina*). *Methods Ecol. Evol.* 12: 311-21.
- Jackson, S. T., et J. T. Overpeck. 2000. Responses of Plant Populations and Communities to Environmental Changes of the Late Quaternary. *Paleobiology* 26 (4): 194-220.
- Jiménez, L., et J. Soberón. 2022. Estimating the Fundamental Niche: Accounting for the Uneven Availability of Existing Climates in the Calibration Area. *Ecol. Modell.* 464: 109823.
- Jiménez, L., J. Soberón, J. A. Christen, et D. Soto. 2019. On the Problem of Modeling a Fundamental Niche from Occurrence Data. *Ecol. Modell.* 397: 74-83.

- Kadmon, R., O. Farber, et A. Danin. 2004. Effect of Roadside Bias on the Accuracy of Predictive Maps Produced by Bioclimatic Models. *Ecol. Appl.* 14 (2): 401-13.
- Kageyama, M., N. Combourieu Nebout, P. Sepulchre, O. Peyron, G. Krinner, G. Ramstein, et J.-P. Cazet. 2005. The Last Glacial Maximum and Heinrich Event 1 in Terms of Climate and Vegetation around the Alboran Sea: A Preliminary Model-Data Comparison. *CR GEOSCI.* 337 (10-11): 983-92.
- Karger, D. N., O. Conrad, J. Bönher, T. Kawohl, H. Kreft, R. W. Soria-Auza, N. E. Zimmermann, H. P. Linder, et M. Kessler. 2017. Climatologies at high resolution for the earth's land surface areas. *Sci. Data* 4: 170122.
- Kass, J. M., R. Muscarella, P. J. Galante, C. L. Bohl, G. E. Pinilla-Buitrago, R. A. Boria, M. Soley-Guardia, et R. P. Anderson. 2021. ENMeval 2.0: Redesigned for Customizable and Reproducible Modeling of Species' Niches and Distributions. *Methods Ecol. Evol.* 12: 1602-1608.
- Latombe, G., A. Burke, M. Vrac, G. Levvasseur, C. Dumas, M. Kageyama, et G. Ramstein. 2018. Comparison of Spatial Downscaling Methods of General Circulation Model Results to Study Climate Variability during the Last Glacial Maximum. *Geosci. Model Dev.* 11 (7): 2563-79.
- Lobo, J. M., A. Jiménez-Valverde, et R. Real. 2008. AUC: A Misleading Measure of the Performance of Predictive Distribution Models. *Glo. Ecol. Biogeogr.* 17 (2): 145-51.
- Lobo, J. M., A. Jiménez-Valverde, et J. Hortal. 2010. The Uncertain Nature of Absences and Their Importance in Species Distribution Modelling. *Ecography* 33: 103-14.
- Machado-Stredel, F., M. E. Cobos, et A. T. Peterson. 2021. A simulation-based method for selecting calibration areas for ecological niche models and species distribution models. *Front. Biogeogr.* 13 (4): e48814.
- Mahalanobis, P.C. 1936. On the Generalized Distance in Statistics. *J. Asiatic Soc. Bengal* 26: 541-88.
- Mammola, Stefano. 2019. Assessing Similarity of N-Dimensional Hypervolumes: Which Metric to Use? *J. of Biogeogr.* 46 (9): 2012-23.
- Marques, R., R. F. Krüger, A. T. Peterson, L. F. de Melo, N. Vicenzi, et D. Jiménez-García. 2020. Climate Change Implications for the Distribution of the Babesiosis and Anaplasmosis Tick Vector, *Rhipicephalus (Boophilus) microplus*. *Vet. Res.* 51 (1): 81.
- Marques, R., R. F. Krüger, S. K. Cunha, A. S. Silveira, D. M.C.C. Alves, G. D. Rodrigues, A. T. Peterson, et D. Jiménez-García. 2021. Climate Change Impacts on *Anopheles (K.) cruzii* in Urban Areas of Atlantic Forest of Brazil: Challenges for Malaria Diseases. *Acta Trop.* 224: 106123.
- Merow, C., M. J. Smith, T. C. Edwards Jr, A. Guisan, S. M. McMahon, S. Normand, W. Thuiller, R. O. Wüest, N. E. Zimmermann, et J. Elith. 2014. What Do We Gain from Simplicity versus Complexity in Species Distribution Models? *Ecography* 37 (12): 1267-81.
- Merow, C., M. J. Smith, et J. A. Silander. 2013. A Practical Guide to MaxEnt for Modeling Species' Distributions: What It Does, and Why Inputs and Settings Matter. *Ecography* 36 (10): 1058-69.
- Morales, N. S., I. C. Fernández, et V. Baca-González. 2017. MaxEnt's Parameter Configuration and Small Samples: Are We Paying Attention to Recommendations? A Systematic Review. *PeerJ* 5: e3093.
- Muscarella, R., P. J. Galante, M. Soley-Guardia, R. A. Boria, J. M. Kass, M. Uriarte, et R. P. Anderson. 2014. ENMeval: An R Package for Conducting Spatially Independent Evaluations and Estimating Optimal Model Complexity for MAXENT Ecological Niche Models. *Methods Ecol. Evol.* 5 (11): 1198-1205.
- Myers, C. E., A. L. Stigall, et B. S. Lieberman. 2015. PaleoENM: Applying Ecological Niche Modeling to the Fossil Record. *Paleobiology* 41 (2): 226-44.
- Nakazawa, Y., A. T. Peterson, E. Martínez-Meyer, et A. G. Navarro-Sigüenza. 2004. Seasonal Niches of Neartic-Neotropical Migratory Birds: Implications for the Evolution of Migration. *Auk* 121 (2): 610.
- Nuñez-Penichet, C., L. Osorio-Olvera, V. H. Gonzalez, M. E. Cobos, L. Jiménez, D. A. DeRaad, A. Alkische, et al. 2021a. Geographic Potential of the World's Largest Hornet, *Vespa mandarinia* Smith (Hymenoptera: Vespidae), Worldwide and Particularly in North America. *PeerJ* 9: e10690.
- Nuñez-Penichet, C., M. E. Cobos, et J. Soberón. 2021b. Non-Overlapping Climatic Niches and Biogeographic Barriers Explain Disjunct Distributions of Continental *Urania* Moths. *Front. Biogeogr.* 13 (2): e5214.
- Owens, H. L., L. P. Campbell, L. L. Dornak, E. E. Saupe, N. Barve, J. Soberón, K. Ingenloff, et al. 2013. Constraints on Interpretation of Ecological Niche Models by Limited Environmental Ranges on Calibration Areas. *Ecol. Modell.* 263: 10-18.
- Papeş, M., et P. Gaubert. 2007. Modelling Ecological Niches from Low Numbers of Occurrences: Assessment of the Conservation Status of Poorly Known Viverrids (Mammalia, Carnivora) across Two Continents: Ecological Niche Modelling of Poorly Known Viverrids. *Divers. Distrib.* 13 (6): 890-902.
- Pearson, R. G., et T. P. Dawson. 2003. Predicting the Impacts of Climate Change on the Distribution of Species: Are Bioclimate Envelope Models Useful? *Glob. Ecol. Biogeogr.* 12 (5): 361-71.
- Pearson, R. G., C. J. Raxworthy, M. Nakamura, et A. T. Peterson. 2006. Predicting Species Distributions from Small Numbers of Occurrence Records: A Test Case Using Cryptic Geckos in Madagascar. *J. Biogeogr.* 34 (1): 102-17.
- Peterson, A. T., et Y. Nakazawa. 2007. Environmental Data Sets Matter in Ecological Niche Modelling: An Example with *Solenopsis invicta* and *Solenopsis richteri*. *Glo. Ecol. Biogeogr.* 17: 135-144.
- Peterson, A. T., et A. G. Navarro-Sigüenza. 2009. Making Biodiversity Discovery More Efficient: An Exploratory Test Using Mexican Birds. *Zootaxa* 2246 (1): 58-66.
- Peterson, A. T., et J. Soberón. 2012. Species Distribution Modeling and Ecological Niche Modeling: Getting the Concepts Right. *Nat. Conserv.* 10 (2): 102-7.

- Peterson, A. T., M. Papeş, et M. Eaton. 2007. Transferability and Model Evaluation in Ecological Niche Modeling: A Comparison of GARP and Maxent. *Ecography* 30 (4): 550-60.
- Peterson, A. T., M. Papeş, et J. Soberón. 2008. Rethinking Receiver Operating Characteristic Analysis Applications in Ecological Niche Modeling. *Ecol. Modell.* 213 (1): 63-72.
- Peterson, A. T., J. Soberón, R. G. Pearson, R. P. Anderson, E. Martinez-Meyer, M. Nakamura, et M. B. Araújo. 2011. *Ecological Niches and Geographic Distributions*. Princeton University Press, Princeton.
- Peterson, A. T., M. E. Cobos, et D. Jiménez-García. 2018. Major Challenges for Correlational Ecological Niche Model Projections to Future Climate Conditions: Climate Change, Ecological Niche Models, and Uncertainty. *Ann. N. Y. Acad. Sci.* 1429: 66-77.
- Phillips, S. J. 2021. maxnet: Fitting “Maxent” Species Distribution Models with “glmnet”. R package version 0.1.4, available at <https://cran.r-project.org/web/packages/maxnet/maxnet.pdf>
- Phillips, S. J., et M. Dudík. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31 (2): 161-75.
- Phillips, S. J., R. P. Anderson, et R. E. Schapire. 2006. Maximum Entropy Modeling of Species Geographic Distributions. *Ecol. Modell.* 190 (3-4): 231-59.
- Phillips, S. J., R. P. Anderson, M. Dudík, R. E. Schapire, et M. E. Blair. 2017. Opening the Black Box: An Open-Source Release of Maxent. *Ecography* 40 (7): 887-93.
- Pierrat, B. 2011. *Macroécologie des échinides de l’océan Austral: Distribution, Biogéographie et Modélisation*. Ph.D thesis, Université de Bourgogne, Dijon (France).
- Pocheville, A. 2015. *The Ecological Niche: History and Recent Controversies*. In *Handbook of Evolutionary Thinking in the Sciences*, édité par T. Heams, Philippe Huneman, G. Lecointre, et Marc Silberstein, 547-86. Dordrecht: Springer Netherlands.
- Proosdij, A. S. J., M. S. M. Sosef, J. J. Wieringa, et N. Raes. 2016. Minimum Required Number of Specimen Records to Develop Accurate Species Distribution Models. *Ecography* 39 (6): 542-52.
- Pulliam, H.R. 2000. On the Relationship between Niche and Distribution. *Ecol. Lett.* 3 (4): 349-61.
- Qiao, H., J. Soberón, et A. T. Peterson. 2015. No Silver Bullets in Correlative Ecological Niche Modelling: Insights from Testing among Many Potential Algorithms for Niche Estimation. *Methods Ecol. Evol.* 6 (10): 1126-36.
- Qiao, H., A. T. Peterson, L. P. Campbell, J. Soberón, L. Ji, et L. E. Escobar. 2016. NicheA: Creating Virtual Species and Ecological Niches in Multivariate Environmental Scenarios. *Ecography* 39 (8): 805-13.
- Radosavljevic, A., et R. P. Anderson. 2014. Making Better Maxent Models of Species Distributions: Complexity, Overfitting and Evaluation. *J. of Biogeogr.* 41 (4): 629-43.
- Randin, C. F., T. Dirnböck, S. Dullinger, N. E. Zimmermann, M. Zappa, et A. Guisan. 2006. Are Niche-Based Species Distribution Models Transferable in Space? *J. of Biogeogr.* 33 (10): 1689-1703.
- Raxworthy, C. J., E. Martinez-Meyer, N. Horning, R. A. Nussbaum, G. E. Schneider, M. A. Ortega-Huerta, et A. T. Peterson. 2003. Predicting Distributions of Known and Unknown Reptile Species in Madagascar. *Nat.* 426 (6968): 837-41.
- Rödder, D., et J. O. Engler. 2011. Quantitative Metrics of Overlaps in Grinnellian Niches: Advances and Possible Drawbacks: Quantitative Metrics of Niche Overlap. *Glob. Ecol. Biogeogr.* 20 (6): 915-27.
- Roberts, D. R., V. Bahn, S. Ciuti, M. S. Boyce, J. Elith, G. Guillera-Arroita, S. Hauenstein, et al. 2017. Cross-Validation Strategies for Data with Temporal, Spatial, Hierarchical, or Phylogenetic Structure. *Ecography* 40 (8): 913-29.
- Roche, D. M., C. Dumas, M. Bügelmayer, S. Charbit, et C. Ritz. 2014. Adding a Dynamical Cryosphere to ILOVECLIM (Version 1.0): Coupling with the GRISLI Ice-Sheet Model. *Geosci. Model Dev.* 7 (4): 1377-94.
- Royle, J. A., et R. M. Dorazio. 2006. Hierarchical Models of Animal Abundance and Occurrence. *J. Agric. Biol. Environ. Stat.* 11 (3): 249-63.
- Saupe, E.E., V. Barve, C.E. Myers, J. Soberón, N. Barve, C.M. Hensz, A.T. Peterson, H.L. Owens, et A. Lira-Noriega. 2012. Variation in Niche and Distribution Model Performance: The Need for *a Priori* Assessment of Key Causal Factors. *Ecol. Modell.* 237-238: 11-22.
- Saupe, E. E., C. E. Myers, A. T. Peterson, J. Soberón, J. Singarayer, P. Valdes, et H. Qiao. 2019. Spatio-Temporal Climate Change Contributes to Latitudinal Diversity Gradients. *Nat. Ecol. Evol.* 3 (10): 1419-29.
- Sbrocco, E. J., et P. H. Barber. 2013. MARSPEC: Ocean Climate Layers for Marine Spatial Ecology: Ecological Archives E094-086. *Ecology* 94 (4): 979-979.
- Shecheglovitova, M., et R. P. Anderson. 2013. Estimating Optimal Complexity for Ecological Niche Models: A Jackknife Approach for Species with Small Sample Sizes. *Ecol. Modell.* 269: 9-17.
- Siddall, M., E.J. Rohling, A. Almogi-Labin, Ch. Hemleben, D. Meischner, I. Schmelzer, et D.A. Smeed. 2003. Sea-Level Fluctuations during the Last Glacial Cycle. *Nature* 423: 853-58.
- Signor, P. W., et J. H. Lipps. 1982. Gradual Extinction Patterns and Catastrophes in the Fossil Record. *Geol. Soc. Am. Bull.* 190: 291-96.
- Sillero, N. 2011. What Does Ecological Modelling Model? A Proposed Classification of Ecological Niche Models Based on Their Underlying Methods. *Ecol. Modell.* 222 (8): 1343-46.
- Sillero, N., et A. M. Barbosa. 2020. Common Mistakes in Ecological Niche Models. *Int J Geogr. Inf. Sci.* 35 (2): 213-226.
- Sillero, Neftalí, Salvador Arenas-Castro, Urtzi Enriquez-Urzelai, Cândida Gomes Vale, Diana Sousa-Guedes, Fernando Martínez-Freiria, Raimundo Real, et A.M. Barbosa. 2021.

- Want to Model a Species Niche? A Step-by-Step Guideline on Correlative Ecological Niche Modelling. *Ecol. Modell.* 456: 109671.
- Singarayer, J. S., et P. J. Valdes. 2010. High-latitude climate sensitivity to ice-sheet forcing over the last 120 kyr. *Quat. Sci. Rev.* 29: 43-55.
- de Siqueira, M. F., G. Durigan, P. de Marco Jr, et A. T. Peterson. 2009. Something from Nothing: Using Landscape Similarity and Ecological Niche Modeling to Find Rare Plant Species. *J. Nat. Conserv.* 17 (1): 25-32.
- Soberón, J. 2007. Grinnellian and Eltonian Niches and Geographic Distributions of Species. *Ecol. Lett.* 10 (12): 1115-23.
- Soberón, J. M. 2010. Niche and Area of Distribution Modeling: A Population Ecology Perspective. *Ecography* 33 (1): 159-67.
- Soberón, J., et B. Arroyo-Peña. 2017. Are Fundamental Niches Larger than the Realized? Testing a 50-Year-Old Prediction by Hutchinson. *PLoS ONE* 12 (4): e0175138.
- Soberón, J., et M. Nakamura. 2009. Niches and Distributional Areas: Concepts, Methods, and Assumptions. *Proc. Natl. Acad. Sci U. S. A.* 106 (2): 19644-50.
- Soberón, J., et A. T. Peterson. 2005. Interpretation of Models of Fundamental Ecological Niches and Species' Distributional Areas. *Biodiv. Inform.* 2: 1-10.
- . 2011. Ecological Niche Shifts and Environmental Space Anisotropy: A Cautionary Note. *Rev. Mex. Biodivers.*, n° 82: 1348-55.
- . 2020. What Is the Shape of the Fundamental Grinnellian Niche? *Theor. Ecol.* 13 (1): 105-15.
- Sobral-Souza, T., J. Pereira Santos, M. E. Maldaner, M. S. Lima-Ribeiro, et M. C. Ribeiro. 2021. EcoLand: A Multiscale Niche Modelling Framework to Improve Predictions on Biodiversity and Conservation. *Perspect. Ecol. Conserv.* 19 (3): 362-68.
- Sohn, N., M. H. Fernandez, M. Papes, et M. Anciães. 2013. Ecological Niche Modeling in Practice: Flagship Species and Regional Conservation Planning. *Oecologia Aust.* 17 (3): 429-40.
- Stockwell, D. R.B., et I. R. Noble. 1992. Induction of Sets of Rules from Animal Distribution Data: A Robust and Informative Method of Data Analysis. *Math. Comput. Simul.* 33 (5-6): 385-90.
- Sweeney, A. W., N. W. Beebe, R. D. Cooper, J. T. Bauer, et A. T. Peterson. 2006. Environmental Factors Associated with Distribution and Range Limits of Malaria Vector *Anopheles farauti* in Australia. *J. Med. Entomol.* 43 (5): 1068-75.
- Vaissié, E. 2021. Mobility of Paleolithic Populations: Biomechanical Considerations and Spatiotemporal Modelling. *PaleoAnthropology* 1: 120–144.
- Valavi, R., J. Elith, J. J Lahoz-Monfort, et G. Guillera-Arroita. 2018. BlockCV: An r Package for Generating Spatially or Environmentally Separated Folds for K-fold Cross-validation of Species Distribution Models. *Methods in Ecol. Evol.* 10 (2): 225-32.
- Valdes, P. J., E. Armstrong, M. P. S. Badger, C. D. Bradshaw, F. Bragg, M. Crucifix, T. Davies-Barnard, et al. 2017. The BRIDGE HadCM3 Family of Climate Models: HadCM3@Bristol v1.0. *Geosci. Model Dev.* 10 (10): 3715-43.
- Van Aelst, Stefan, et Peter Rousseeuw. 2009. Minimum Volume Ellipsoid. *Wiley Interdiscip. Rev. Comput. Stat.* 1 (1): 71-82.
- VanDerWal, J., L. P. Shoo, C. Graham, et S. E. Williams. 2009. Selecting Pseudo-Absence Data for Presence-Only Distribution Modeling: How Far Should You Stray from What You Know? *Ecol. Modell.* 220 (4): 589-94.
- Varela, S., R. P. Anderson, R. García-Valdés, et F. Fernández-González. 2014. Environmental Filters Reduce the Effects of Sampling Bias and Improve Predictions of Ecological Niche Models. *Ecography* 37: 1084-91.
- Varela, S., M. S. Lima-Ribeiro, et L. C. Terribile. 2015. A Short Guide to the Climatic Variables of the Last Glacial Maximum for Biogeographers. *PLoS ONE* 10 (6): e0129037.
- Vignali, S., A. G. Barras, R. Arlettaz, et V. Braunisch. 2020. SDMtune: An R Package to Tune and Evaluate Species Distribution Models. *Ecol. Evol.* 10 (20): 11488-506.
- Vignoles, A. 2021. Trajectoires technologiques et dynamiques de niches éco-culturelles du Gravettien moyen au Gravettien récent en France. Ph.D thesis, Université de Bordeaux, pessac (France).
- Vignoles, A., W. E. Banks, L. Klaric, M. Kageyama, M. E Cobos, et D. Romero-Alvarez. 2021. Investigating Relationships between Technological Variability and Ecology in the Middle Gravettian (ca. 32-28 Ky Cal. BP) in France. *Quat. Sci. Rev.* 253: 106766.
- Volterra, V. 1926. Variazioni e fluttuazioni del numero d'individui in specie animali conviventi. *Accademia Nazionale dei Lincei, Roma (Italia).*
- Vrac, M., P. Marbaix, D. Paillard, et P. Naveau. 2007. Non-Linear Statistical Downscaling of Present and LGM Precipitation and Temperatures over Europe. *Clim. Past* 3: 669-82.
- Walker, P. A., et K. D. Cocks. 1991. HABITAT: A Procedure for Modelling a Disjoint Environmental Envelope for a Plant or Animal Species. *Glob. Ecol. Biogeogr.* 1 (4): 108-118.
- Warren, D. L. 2012. In Defense of 'Niche Modeling'. *Trends Ecol. Evol.* 27 (9): 497-500.
- Warren, D. L., et S. N. Seifert. 2011. Ecological Niche Modeling in Maxent: The Importance of Model Complexity and the Performance of Model Selection Criteria. *Ecol. Appl.* 21 (2): 335-42.
- Warren, D. L., R. E. Glor, et M. Turelli. 2008. Environmental Niche Equivalency versus Conservatism: Quantitative Approaches to Niche Evolution. *Evol.* 62 (11): 2868-83.
- Warren, D. L., A. N. Wright, S. N. Seifert, et H. B. Shaffer. 2014. Incorporating Model Complexity and Spatial Sampling Bias into Ecological Niche Models of Climate Change Risks Faced by 90 California Vertebrate Species of Concern. *Divers. Distrib.* 20 (3): 334-43.
- Warren, D. L., A. Dornburg, K. Zapfe, et T. L. Iglesias. 2021a. The Effects of Climate Change on Australia's Only Endemic Pokémon: Measuring Bias in Species Distribution Models. *Methods Ecol. Evol.* 12 (6): 985-995.



- Warren, D. L., N. J. Matzke, M. Cardillo, J. B. Baumgartner, L. J. Beaumont, M. Turelli, R. E. Glor, et al. 2021b. ENMTools 1.0: An R Package for Comparative Ecological Biogeography. *Ecography* 44: ecog.05485.
- Whittaker, R. H., S. A. Levin, et R. B. Root. 1973. Niche, Habitat, and Ecotope. *Am. Nat.* 107 (955): 321-38.
- Wieczorek, J., Q. Guo, et R. Hijmans. 2004. The Point-Radius Method for Georeferencing Locality Descriptions and Calculating Associated Uncertainty. *Int. J. Geogr. Inf. Sc.* 18 (8): 745-67.
- Williams, J. W., et S. T. Jackson. 2007. Novel Climates, No-Analog Communities, and Ecological Surprises. *Front. Ecol. Environ.* 5 (9): 475-82.
- Williams, J. W., S. T. Jackson, et J. E. Kutzbach. 2007. Projected Distributions of Novel and Disappearing Climates by 2100 AD. *Proc. Natl. Acad. Sc. U. S. A.* 104 (14): 5738-42.
- Wisz, M. S., R. J. Hijmans, J. Li, A. T. Peterson, C. H. Graham, A. Guisan, et NCEAS Predicting Species Distributions Working Group†. 2008. Effects of Sample Size on the Performance of Species Distribution Models. *Divers. Distrib.* 14 (5): 763-73.
- Zhu, G., W. Bu, Y. Gao, et G. Liu. 2012. Potential Geographic Distribution of Brown Marmorated Stink Bug Invasion (*Halyomorpha halys*). *PLoS ONE* 7 (2): e31246.
- Zuquim, G., F. R. C. Costa, H. Tuomisto, G. M. Moulatlet, et F. O. G. Figueiredo. 2020. The Importance of Soils in Predicting the Future of Plant Habitat Suitability in a Tropical Forest. *Plant Soil* 450 (1-2): 151-70.
- Zurell, D., J. Elith, et B. Schröder. 2012. Predicting to New Environments: Tools for Visualizing Model Behaviour and Impacts on Mapped Distributions. *Divers. Distrib.* 18 (6): 628-34.