

6- Modèles

Modélisation de la distribution des espèces

Pour déterminer la distribution d'une espèce, on distingue deux grandes catégories de modèles (Carpenter et al., 1993; Zimmermann and Kienast, 1999; Morin et Thuiller.,2009). D'une part, les modèles dynamique de végétation appelés aussi les modèles mécanistiques (Kearney et al., 2008; Kearney et Porter, 2009; Evans et al., 2012.) et d'autre part, les modèles empiriques, dite statistiques (Carpenter et al., 1993; Zimmermann and Kienast, 1999).

Les modèles dynamique de végétation ont pour objectif simulé la performance d'une espèce dans du temps ou selon les conditions environnementales qui varient dans l'espaces. ils sont basés sur les spécificités physiologiques des espèces tandis que les modèles empériques sont fondés sur la recherche de relations de corrélation statistiques entre les variations spatiale des conditions environnementales (climatiques, topographiques,...) et les distributions observées des espèces (Guisan and Zimmermann, 2000).

5.1. Modèles mécanistes

Les modèles mécanistes permettent la simulation et l'évolution spatiale de traits physiologiques selon un gradient écologique. L'ensemble des prédictions sur performances vis -à-vis chacun des variables utilisées, permet la visualisation, dans une zone géographique les limites de niche fondamentale et les caractéristiques comportementales et physiologiques des organismes, ce qui permet d'expliquer les limites de répartition des espèces (Kearney et Porter, 2009).

Le modèle mécaniste repose sur des origines biophysiques qui permettent le transfert d'un espace environnemental à un ensemble de prouesses physiologiques des espèces. Il nécessite des connaissances approfondies sur la physiologie des espèces. Elle permet aussi d'évaluées degré d'équilibre des organismes avec leur environnement (Kearney et al., 2008).

Toutefois, cette approche est lourde et son succès est conditionné essentiellement par le niveau de connaissance de la physiologie des espèces (Kearney et Porter, 2004; Kearney et Porter, 2009 ; Dormann et al., 2012).

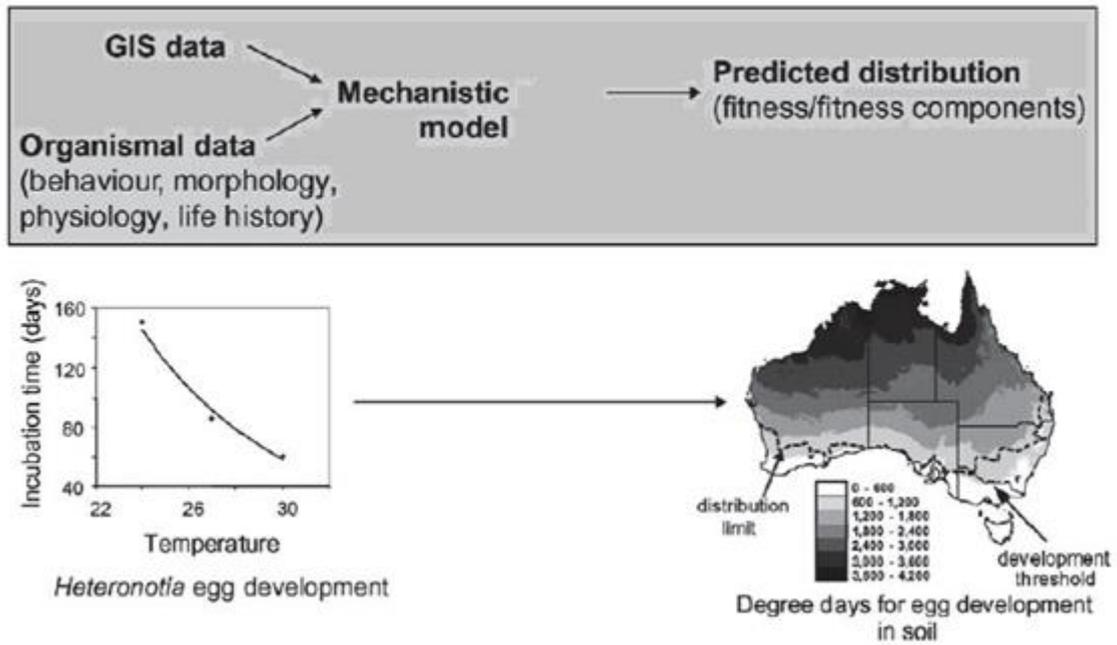


Figure 1: Schéma de l'approche mécaniste de modélisation

5.2. Modèles corrélatifs

Les méthodes empiriques ou corrélatives, appelées aussi modèles statistiques, se sont des modèles basés sur les niches écologique « Niche-based-models » permet de déterminé directement les relations entre la présence de l'espèces et son environnement pour identifier et quantifier statistiquement les processus qui conditionnent la distribution de l'espèces. néanmoins ce type de modèle exigent d'une part, des données sur l'occurrence spatiale des espèces et d'autre part des données environnementales spatialisées.

L'approche corrélative ne nécessite pas une connaissance de l'autécologie des espèces, largement utilisée et favorisé par la croissante évolution des modèles statistiques de plus en plus performants, aussi la disponibilité des base de données à référence spatiale d'occurrence de la flore, de données des facteurs écologiques spatialisées et de Systèmes d'Information Géographiques (SIG) (Bertrand.,2012).

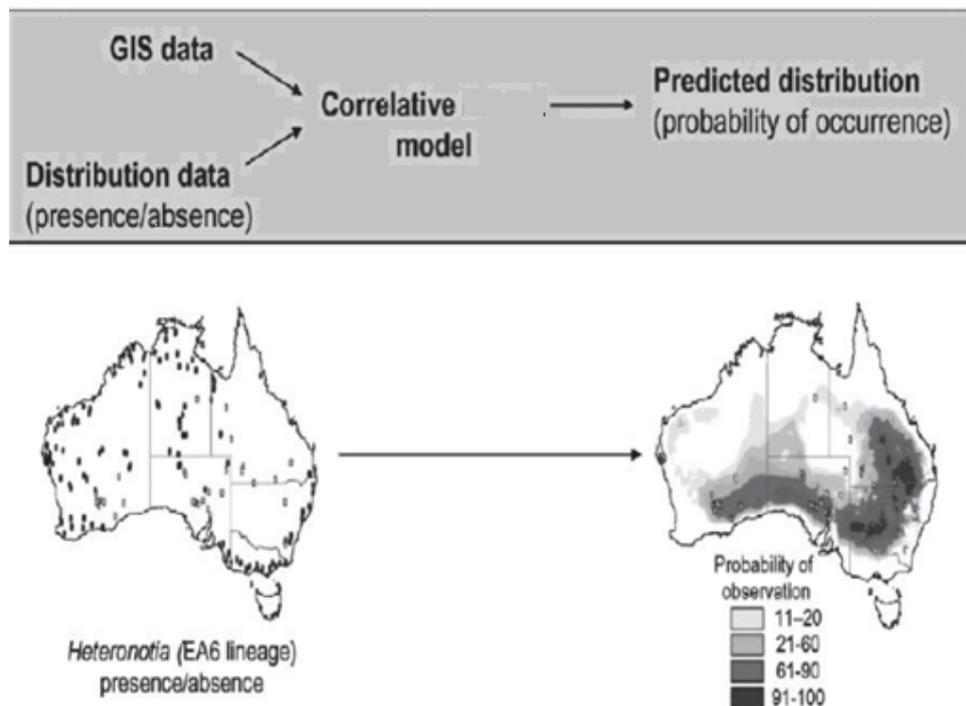


Figure 2: Approche corrélative de prédiction de la distribution d'une espèce de gecko australien (*Heteronotiabinoei*).

5.2.1 Modèles SDMs

Appelé *Species Distribution Model* (SDM). Ces derniers combinent les données de l'espèce avec des paramètres environnementaux pour générer des cartes de qualité d'habitat suivant différentes méthodes. Ils sont une approche alternative qui permet d'utiliser efficacement les observations et prouver ensuite les interprétations ainsi que la compréhension de la distribution des espèces (Walaa., 2012).

6- Application SDM

Méthodes de Maximum Entropie (MaxEnt)

MaxEnt est un programme de classification probabiliste linéaire basé sur l'entropie maximale et estime la probabilité de présence d'espèces en fonction des variables environnementales. Il ne nécessite que les données concernant la présence d'espèces (Phillips et al., 2006).

L'application du formalisme de l'entropie maximale à la modélisation de la distribution des espèces a été introduite par Phillips et al. (2004). L'Approches de modélisation MaxEnt

compare les données de présence observées ($y = 1$) dans un espace écologique dans une région d'étude (z est un vecteur de prédictors environnementaux.), $f(z)$ est définie comme la densité de probabilité des prédictors dans l'espace disponible et $f_1(z)$ comme la densité de probabilité des multi-variées des variables environnementales et les emplacements dans la région où l'espèce se trouve (Elith et al. 2011). MaxEnt utilise les prédictors de l'occurrence et l'échantillon d'occurrence utilisé pour estimer le rapport $f_1(z) / f(z)$. L'algorithme d'optimisation cherche $f_1(z)$ qui minimise la distance de $f(z)$. Le but de la méthode est de modéliser la probabilité de présence de l'espèce, selon les conditions environnementales disponibles (Gomes et al., 2018).

En effet, $f(z)$ est considéré comme un modèle nul pour $f_1(z)$ puisqu'il n'y a aucune raison de s'attendre à ce que l'espèce préfère des conditions environnementales particulières en l'absence de données d'occurrence. Dans ce dernier cas, la meilleure prédiction est que l'espèce occupe des conditions environnementales proportionnellement à sa disponibilité dans la région. Dans MaxEnt, cette distance de $f(z)$ est considérée comme l'entropie relative de $f_1(z)$ par rapport à $f(z)$. Puisque MaxEnt ne modélise pas spécifiquement les données de présence mais plutôt la densité des conditions environnementales utilisées, les sorties brutes de MaxEnt sont ensuite retransformées en sorties logistiques pour être directement interprétées comme une probabilité d'occurrence (Elith et al. 2011).

5.3. Application du modèle

Le principe de la méthode de l'entropie maximale est d'estimer la véritable occurrence d'une espèce, représentée comme une probabilité de présence variant entre 0 et 1. Ainsi, lorsque la probabilité est égale à 1, la présence est estimée comme vraie (Phillips et Dudík., 2008). L'approche utilisée consiste à diviser de manière aléatoire toutes les données en données de "formation" et données "test" afin d'obtenir des données indépendantes pour évaluer la performance du modèle (Fielding et Bell., 1997 ; Guisan et al., 2003). Ainsi, un total de 75% des données d'occurrence sélectionnées au hasard a été utilisé pour la modélisation (données formation) et les 25% restants (données test) sont utilisés pour valider le modèle, pour 10 répétitions dont la sélection est différente pour chaque répétition afin d'obtenir une estimation moyenne de la performance du modèle (Phillips et al., 2006 ; Aguirre et al., 2013 ; Qin et al., 2017).

5.4. Performance et évaluation du modèle

L'évaluation des performances est une étape importante dans la validation du modèle. L'analyse des caractéristiques de fonctionnement récepteur ROC (*receiver operating characteristic*) (Fielding & Belle., 1997, Phillips et al., 2009) permet l'évaluation de la performance du modèle à travers la valeur de l'AUC (*Area Under Curve*).

L'AUC d'un ROC est actuellement utilisé comme paramètre de discrimination la plus couramment utilisée pour l'évaluation de la performance de la modélisation de la distribution des espèces (Swets, 1988, Fielding et Bell, 1997). Elle traduit la capacité des modèles à distinguer les présences et les absences des espèces. Elle intègre toutes les seuils de tolérance de l'espèce aux paramètres explicatives, dans une seule mesure déterminante de la qualité et la performance de la simulation avec une illustration graphique (Guisan et al., 2017). AUC défini le succès de la prédiction du SDM, elle exprime le pourcentage de présence et absence d'une espèce dans un échantillon bien prédite par le modèle (Bertrand et al., 2012) Une seule valeur quantitative pour caractériser la performance du modèle.

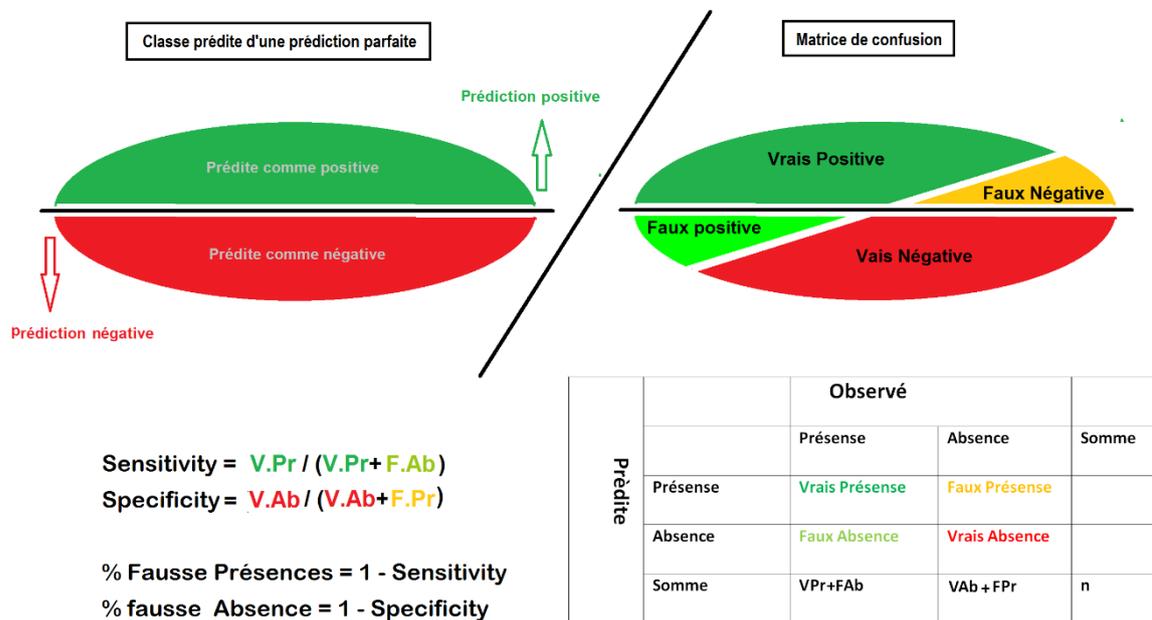


Figure 23: Matrice de confusion d'erreur de prédiction

La première étape consiste à calculer la sensibilité (Sensitivity); défini par le taux de présences correctement prédit et la spécificité (Specificity); défini par le pourcentage d'absences correctement prédit (Manel et al., 2001 ; Liu et al., 2011) (Figure 23).

La seconde étape consiste à tracer sur le graphique les valeurs de sensibilité par rapport à la valeur correspondante de $[1 - \text{spécificité}]$, qui exprime le pourcentage fausses prédictions d'absences (taux de commission). Ainsi la courbe caractéristique de fonctionnement du récepteur (ROC) est établie illustrant une seule valeur quantitative pour caractériser la performance du modèle (Figure 24).

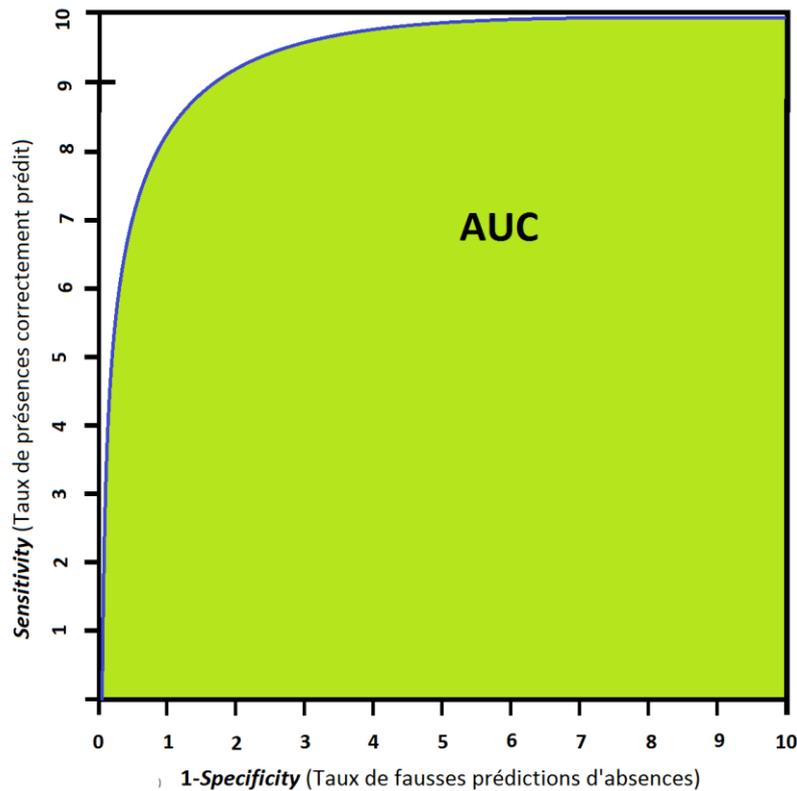


Figure 24: ROC la courbe caractéristique de fonctionnement du récepteur (*the receiver operating characteristic*).

L'analyse des caractéristiques de fonctionnement du récepteur ROC (*receiver operating characteristic*) (Fielding & Belle., 1997, Phillips et al., 2009) permet l'évaluation de la performance du modèle à travers la valeur de l'AUC (*Area Under Curve*).

Cette valeur varie de 0 à 1 ; dont le 1 représente la performance maximale,

Valeur d'AUC entre 0,5 et 0,7 considérée comme faible,

Valeur d'AUC entre 0,7 et 0,9 considérée comme bonne,

Valeur d'AUC supérieure à 0,9 considérée comme haute. (Swets., 1988;Lobo et al., 2008 ; Reddy et al., 2015; Araujo et al., 2015; Guisan et al., 2017).