

Guide méthodologique pour la réalisation de cartes de sensibilité chez la Cistude d'Europe

Note technique



Siège social :

MnHn – CP41
57 Rue Cuvier

Siège administratif :

c/o Isabelle Chauvin
2014 Route de Roquefort
32360 Peyrusse-Massas

contact@lashf.org

<http://lashf.org>



Document réalisé par Audrey Trochet : audrey.trochet@lashf.org

Logiciel utilisés : QGIS, R

Février 2021

Avec le soutien financier de la DREAL Auvergne-Rhône-Alpes



**PRÉFET
DE LA RÉGION
AUVERGNE-RHÔNE-ALPES**

*Liberté
Égalité
Fraternité*

Citation du rapport : Trochet A., (2021). « Guide méthodologique pour la réalisation de cartes de sensibilité chez la Cistude d'Europe – Note technique ». Société Herpétologique de France. 29 p.

Table des matières

Guide méthodologique pour la réalisation de cartes de sensibilité chez la Cistude d'Europe 1

Note technique	1
1. Contexte	4
2. Méthodes utilisées pour la réalisation des cartes de sensibilité	4
2.1. Méthode standard vs modélisation prédictive d'habitats	4
2.2. Principe des modèles prédictifs	5
2.3. Modèle d'entropie maximale (MaxEnt)	6
2.4. Évaluation du modèle prédictif	8
3. Carte de sensibilité nationale de la Cistude d'Europe	8
3.1. Données d'occurrence	8
3.2. Variables environnementales	10
3.3. Carte de sensibilité nationale	11
3.4. Limites du modèle et recommandations	15
4. Bibliographie	16
5. Guide technique pour la réalisation des cartes de sensibilité (méthode MaxEnt)	18
5.1. Données d'occurrence	18
5.1.1. Récolte des données	18
5.1.2. Formatage des données	19
5.2. Variables environnementales	19
5.2.1. Téléchargement des couches habitats	19
5.2.2. Création des rasters « habitats »	20
5.3. Modélisation prédictive d'habitats sous R	26
5.3.1. MaxEnt	26
5.3.2. Évaluation de la robustesse du modèle prédictif	27
5.4. Fichier de sortie raster [QGIS]	29



1. Contexte

La prise en compte des espèces dans le cadre des politiques publiques, et notamment de l'aménagement du territoire, est un enjeu majeur pour la conservation de la biodiversité. Afin de répondre à ce besoin, la mise à disposition d'outils d'aide à la décision, telles que des cartographies d'alerte sur les secteurs de sensibilité pour certaines espèces menacées par exemple, est essentielle. Ces cartographies d'alerte, nommées également « cartes de sensibilité », hiérarchisent les enjeux relatifs à une espèce cible, via la représentation cartographique de zones de vigilance (appelées aussi « zones sensibles ») où l'espèce est vraisemblablement présente, au sein d'un territoire donné. Par conséquent, tout aménagement localisé dans des zones à vigilance élevée devra être évité ou nécessitera une réflexion particulière.

Hormis l'intérêt de tels outils dans l'aménagement du territoire, les cartes de sensibilité permettent également d'orienter les actions de conservation sur l'espèce cible (comme, par exemple, guider les efforts de prospection dans certaines zones pauvres en points d'occurrence, ou localiser des zones favorables susceptibles d'accueillir des renforcements ou des réintroductions de populations) et sont des dispositifs fréquemment utilisés dans les programmes de conservation.

2. Méthodes utilisées pour la réalisation des cartes de sensibilité

2.1. Méthode standard vs modélisation prédictive d'habitats

Les cartes de sensibilité peuvent être réalisées à l'aide de différentes méthodes. La première consiste à représenter les « zones de vigilance » de l'espèce autour des points d'occurrence saisis, à l'aide de zones tampon par exemple, en prenant en compte si possible (ou non) le type d'habitats à proximité. Cette méthode nécessite une excellente connaissance du territoire concerné, et est entièrement dépendante des points d'occurrence disponibles. Afin d'obtenir un jeu de données fortement représentatif de la répartition actuelle de l'espèce cible (et éviter tout biais dans la carte de sensibilité produite), la réalisation d'inventaires exhaustifs peut être préconisée, bien que ces derniers puissent s'avérer coûteux, fastidieux, et très chronophages, notamment à large échelle spatiale.



La seconde méthode passe par de la modélisation prédictive d'habitats, via notamment des approches corrélatives. Bien que la modélisation ne remplace aucunement les connaissances naturalistes, ce type d'outil permet de mieux connaître la distribution potentielle d'une espèce et de ses habitats, et également de mieux appréhender les problématiques de conservation (il conviendra ensuite aux utilisateurs d'amender et valider la carte de sensibilité produite à partir de la modélisation en se rapprochant des experts de l'espèce concernée). La modélisation prédictive d'habitats part du principe que les sites d'observation d'une espèce se trouvent dans des habitats favorables à cette dernière. Le modèle va donc estimer les conditions environnementales favorables à l'espèce à partir des points d'occurrence et des paramètres environnementaux susceptibles d'influencer, entre autres, sa distribution (Pearson *et al.* 2007). Puis, à partir de la relation obtenue, le modèle prédictif va fournir une carte de sensibilité basée sur des probabilités de présence de l'espèce sur l'ensemble du territoire donné (plus de détails ci-dessous). Les « zones de vigilance » pourront donc être associées aux secteurs ayant une forte probabilité de présence prédite par le modèle. Cette méthode, contrairement à la première, peut éventuellement mettre en évidence des zones de vigilance dépourvues de points d'occurrence.

Les deux méthodes présentées ci-dessus sont, la plupart du temps, complémentaires. Pour la réalisation d'une carte de sensibilité à large échelle spatiale (échelle nationale par exemple), il est recommandé de passer via une approche de modélisation pour produire une première carte de sensibilité (du fait de la complexité de mettre en œuvre des inventaires ultra précis à cette échelle), que les utilisateurs seront libres de reproduire à plus petite échelle (régionale ou départementale). Puis, dans un second temps, les cartes de sensibilité produites à l'échelle plus locale pourront être amendées et validées auprès des experts locaux de l'espèce concernée.

2.2. Principe des modèles prédictifs

Pour la réalisation de cartes de sensibilité à partir de modélisation prédictive d'habitats, plusieurs modèles sont disponibles (liste non-exhaustive) : analyse factorielle de niche écologique – ENFA (*Ecological Factor Niche Analysis*), modèle d'entropie maximale – MaxEnt (*Maximum Entropy*) ou modèle linéaire généralisé – GLM (*Generalized Linear Model*). Le principe de la modélisation prédictive d'habitats reste le même pour tous les modèles disponibles, et passe par plusieurs étapes indispensables (Bargain & Fabri 2015 ; Guisan & Zimmermann 2000) :

- Compilation des **points d'occurrence** de l'espèce étudiée sur la zone d'étude
- Sélection, compilation et formatage des **couches de variables environnementales**
- **Réalisation du modèle prédictif** et création de la **carte de sensibilité**



- **Évaluation** de l'efficacité du modèle pour prédire la probabilité d'occurrence de l'espèce étudiée (évaluation de la relation entre les points d'occurrence et les paramètres environnementaux : on parle de significativité de la relation, permettant d'évaluer la pertinence et la fiabilité de la carte produite).

Plusieurs études (Préau *et al.* 2017, 2018) et rapports (Bargain & Fabri 2015 ; Préau *et al.* 2019) ont récemment évalué la fiabilité de chacun des modèles listés ci-dessus. Il s'avère que, dans la majorité des cas, la méthode du modèle d'entropie maximale (**MaxEnt**) apparaît comme étant la plus performante pour modéliser spatialement l'occurrence d'une espèce en fonction de ses données de présence et des variables environnementales caractérisant son habitat (Elith *et al.* 2006 ; Hernandez *et al.* 2006 ; Wang *et al.* 2007). Par conséquent, seule cette méthode sera présentée dans le présent document et utilisée pour la réalisation des cartes de sensibilité nationale de la Cistude d'Europe (*Emys orbicularis*), dans le cadre du PNA II Cistude d'Europe (2020-2029).

2.3. Modèle d'entropie maximale (MaxEnt)

Une des méthodes couramment utilisée pour la réalisation de modèles prédictifs d'habitats (et donc, des cartes de sensibilité) est le modèle d'entropie maximale (ou de maximum d'entropie ; l'entropie correspondant à une caractérisation statistique de l'information manquante ; Venditti 1998). L'approche du maximum d'entropie est une méthode de prédiction spatialisée de l'occurrence d'une espèce basée sur le concept de niche écologique et sur l'utilisation de données de présence (Phillips *et al.* 2006 ; Fig. 1).

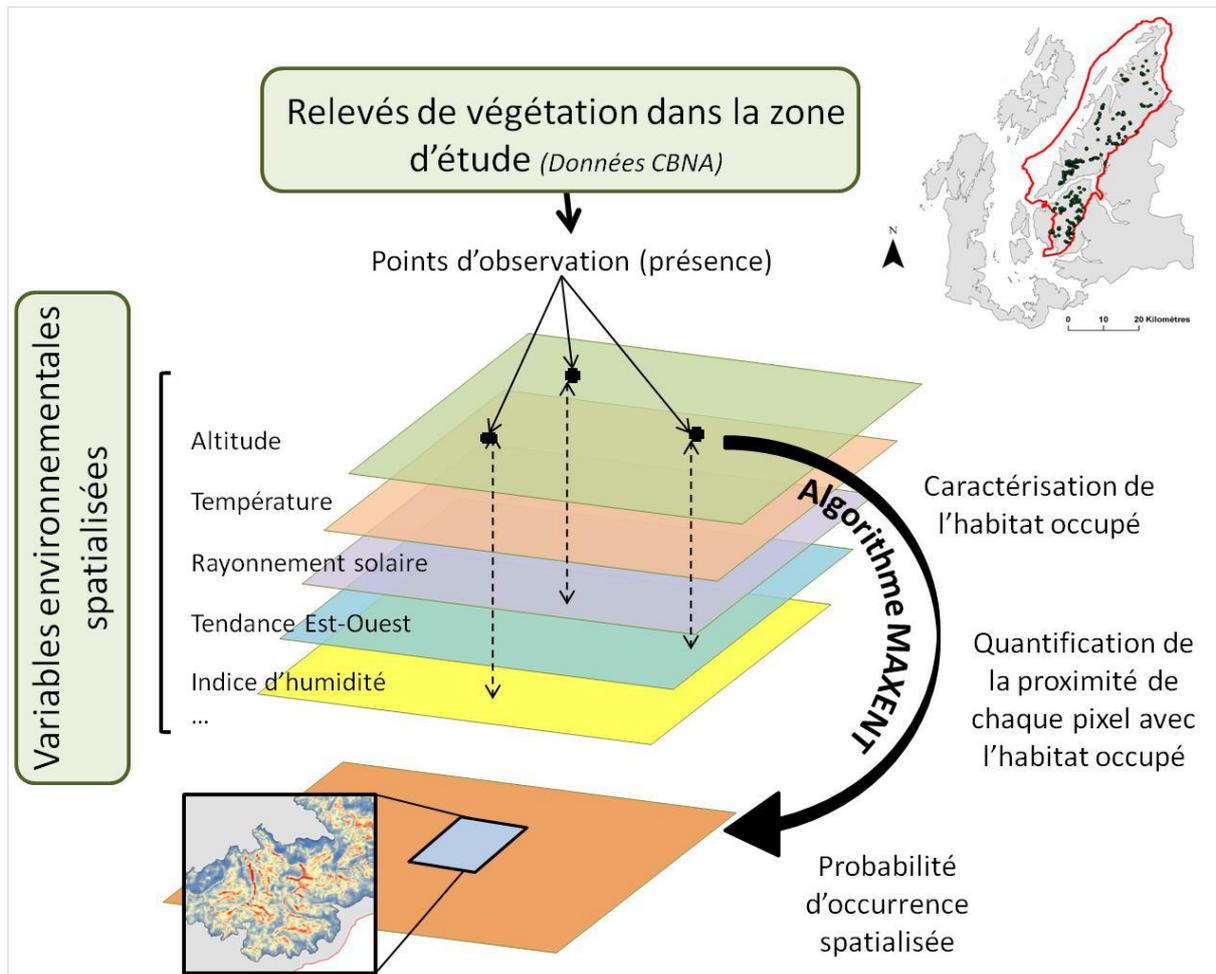


Figure 1. Principe général de la méthode du maximum d'entropie (figure issue de Luque et al. 2012)

Comme les autres modèles prédictifs utilisant des approches corrélatives, le modèle d'entropie maximale (implémenté dans le logiciel MaxEnt, et également disponible sous scripts R) nécessite à la fois des données d'occurrence et des variables environnementales, aussi bien quantitatives que qualitatives (Elith et al. 2011). Notons ici que des données d'absence avérées de l'espèce cible ne sont pas obligatoires pour la mise en œuvre de cette méthode, ce qui apparaît comme étant un avantage du fait que ces données sont souvent difficiles à obtenir (Guisan et al. 2005 ; Pearce & Boyce 2006).

→ Pour en savoir plus sur MaxEnt : https://biodiversityinformatics.amnh.org/open_source/maxent/. Le logiciel MaxEnt (accompagné de tutoriels, en anglais) est disponible gratuitement sur le site.



Toutefois, certaines versions de Windows et Java peuvent contraindre son utilisation. Il est donc préférable d'appliquer cette méthode via l'utilisation de scripts R fournis en fin du présent document.

2.4. Évaluation du modèle prédictif

Une fois le modèle prédictif construit (et la carte de sensibilité obtenue), une validation de ce dernier est nécessaire pour estimer sa robustesse (fiabilité). La pertinence d'un modèle peut être réalisée à l'aide de la méthode de la courbe ROC (*Receiver Operating Characteristic*) et en particulier de l'estimation de l'aire sous cette courbe (AUC : *Area Under Curve*), qui permet de distinguer si les prédictions issues du modèle prédictif sont proches d'une distribution aléatoire, ou si au contraire, les prédictions du modèle sont statistiquement fiables (Luque *et al.* 2012).

La valeur de l'AUC varie entre 0 et 1 (une valeur d'AUC de 0.5 correspond à un modèle aléatoire ; Kato 2012). Plus la valeur de l'AUC est grande, plus le modèle est proche de la réalité. Les valeurs d'AUC, classées en catégories, informent donc sur la performance du modèle prédictif (Araújo *et al.* 2005) : $0.50 < \text{AUC} < 0.60$: modèle prédictif indifférent d'un modèle aléatoire (non représentatif de la réalité) ; $0.60 < \text{AUC} < 0.70$: modèle prédictif médiocre ; $0.70 < \text{AUC} < 0.80$: modèle prédictif moyen ; $0.80 < \text{AUC} < 0.90$: modèle prédictif bon ; $0.90 < \text{AUC} < 1$: modèle prédictif excellent.

3. Carte de sensibilité nationale de la Cistude d'Europe

3.1. Données d'occurrence

Au 31 décembre 2020, une base de données regroupant des données d'occurrence issues de multiples structures (*par ordre alphabétique* : Cistude Nature, Conservatoire d'Espaces Naturels (CEN) de l'Allier, Conservatoire d'Espaces Naturels (CEN) de Corse, École Pratique des Hautes Études (EPHE), Groupe Mammalogique et Herpétologique du Limousin (GMHL), Association Lo Parvi, Nature en Occitanie, Nature Environnement 17 ; Tableau 1) et de synthèses régionale et nationale (GeoNature de la Société Herpétologique de France (SHF), Inventaire national du patrimoine naturel (INPN), Système d'Information sur la Nature et les Paysages (SINP) Occitanie ; Tableau 1) comptabilisait au total **21 171** données d'occurrence de Cistude d'Europe (*Emys orbicularis*). L'ensemble des données se répartit au sein des régions de Nouvelle-Aquitaine, Occitanie, Provence-

Alpes-Côte-D'azur, Auvergne-Rhône-Alpes, Corse et Centre Val de Loire, Bourgogne France-Comté (Fig. 2).

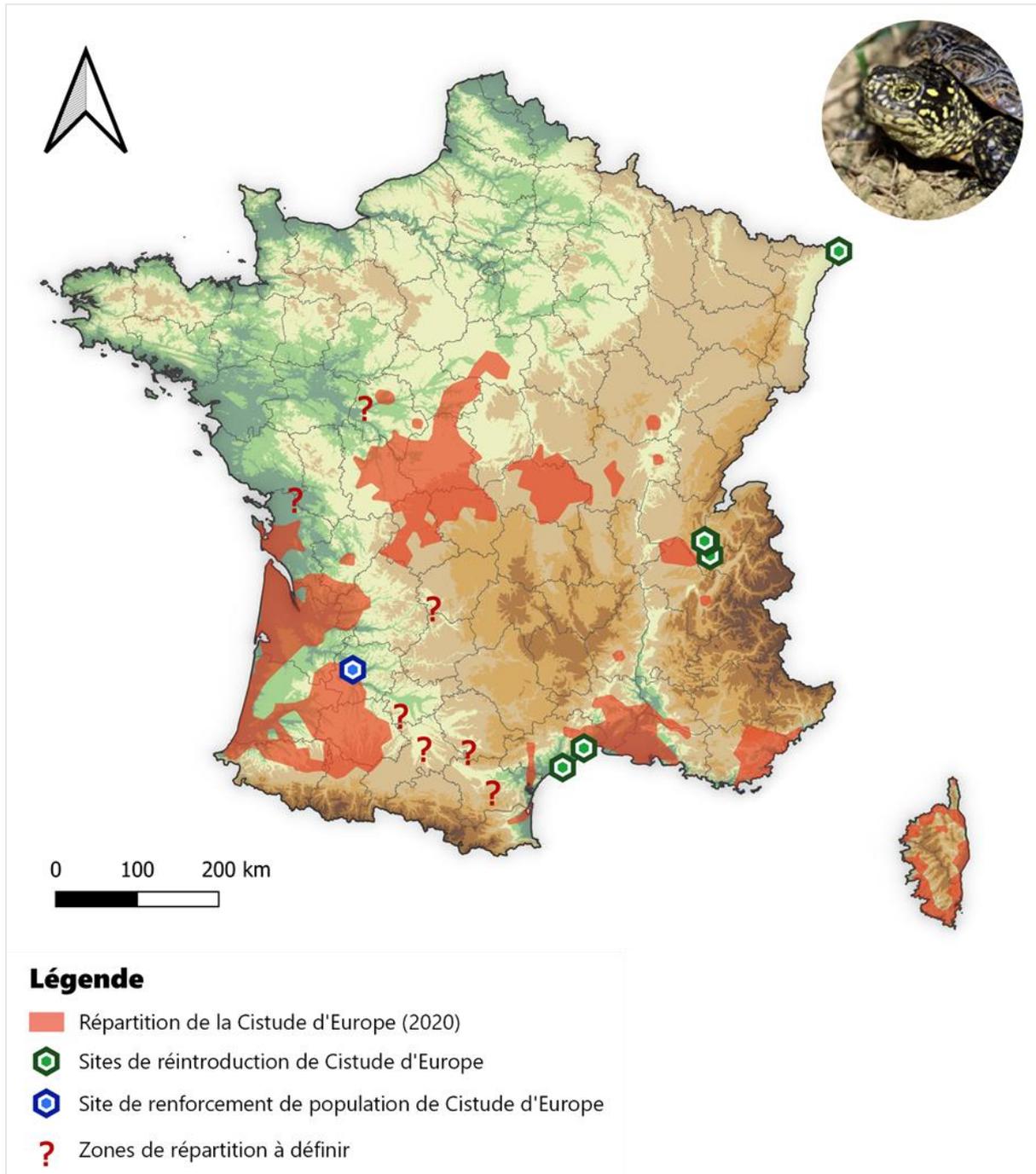




Figure 2. Carte de répartition de la Cistude d'Europe (*Emys orbicularis*) en France métropolitaine (version de décembre 2020)

Tableau 1. Répartition des données d'occurrence de Cistude d'Europe compilées pour la réalisation de la carte de sensibilité (version de décembre 2020 ; CEN : Conservatoire d'Espaces Naturels, GMHL : Groupe Mammalogique et Herpétologique du Limousin, INPN : Inventaire national du patrimoine naturel, SINP : Système d'Information sur la Nature et les Paysages, EPHE : École Pratique des Hautes Études). Les synthèses régionale/nationale avec un * indiquent qu'elles comprennent de multiples structures

Structure	Nombre d'occurrence	Synthèses régionale/nationale	Nombre d'occurrence
CEN Allier	826	GeoNature SHF *	2773
Cistude Nature	393	INPN *	2760
CEN Corse	1328	SINP Occitanie	5221
LO Parvi	1675		
GMHL	371	EPHE	4109
Nature Environnement 17	5824	Nature en Occitanie	1112
		Total	21 171

Les synthèses régionale et nationale (GeoNature SHF, INPN et SINP Occitanie) étant des outils de compilation de données, il convient, avant de réaliser la modélisation sous MaxEnt, de supprimer les éventuels doublons. En effet, des données d'occurrence répétées pourraient biaiser le modèle prédictif en « tirant » les conditions environnementales favorables de l'espèce de manière déséquilibrée. Après un tri de la base de données Cistude d'Europe, **10 198** données d'occurrence non répétées sont gardées et utilisées dans le modèle.

3.2. Variables environnementales

Après discussions avec les experts de l'espèce considérée, les variables environnementales retenues pour la modélisation prédictive d'habitats (et donc, susceptibles d'avoir une influence sur la répartition de la Cistude d'Europe) sont :

- Altitude et pente (BD Alti © IGN)
- *Distance à la plus proche autoroute et à la plus proche route secondaire (Route 500 © IGN)



- *Distance au plus proche réseau ferré (© SNCF)
- *Distance au plus proche cours d'eau et au plus proche point d'eau (BD Carthage © IGN)
- *Distance à la plus proche végétation arbustive, à la plus proche zone agricole, à la plus proche zone urbaine, au plus proche espace ouvert, à la plus proche forêt et à la plus proche prairie (Corine Land Cover)
- Lithologie (BRGM : Bureau de Recherches Géologiques et Minières)
- Précipitation annuelle, rayonnement solaire moyen annuel (1970-2000) et température moyenne annuelle (période 1970-2000) (WorldClim)

Toutes les variables environnementales listées ci-dessus sont implémentées dans le modèle via des fichiers rasters (pixels de 200 m x 200 m). Chaque raster a une résolution de 200 m x 200 m (taille du pixel). Chaque pixel contient une valeur de la couche concernée (valeurs d'altitude ou de température par exemple).

Concernant les cartes indiquant la présence (ou l'absence) d'un habitat spécifique (cartes avec un * dans la liste ci-dessus), tel que les forêts par exemple, les fichiers rasters sont construits à partir du calcul de la distance au plus proche habitat considéré. Ainsi, plutôt que d'attribuer une valeur de 0 (absence de forêt par exemple) ou de 1 (présence de forêt par exemple) dans chaque pixel, une valeur continue (correspondant à la distance à la plus proche forêt par exemple) est associée à chaque pixel. Cette approche permet d'obtenir des couches d'habitats contenant une plus grande variabilité environnementale et ainsi contraindre la relation avec les données d'occurrence de manière plus fine.

→ *Les sources des couches (ainsi que la méthode de préparation des données environnementales pour la modélisation) sont détaillées dans la partie « Guide méthodologique » en bas du présent document.*

3.3. Carte de sensibilité nationale

Une carte de sensibilité nationale de la Cistude d'Europe (*Emys orbicularis*) est obtenue à l'issue de la modélisation sous MaxEnt (Fig. 3).

La valeur d'AUC est de 0.869, correspondant à un modèle prédictif de bonne qualité. Les variables ayant une importance relative élevée (c'est-à-dire celles qui influencent le plus la présence de l'espèce) sont la température moyenne annuelle (27.1% : probabilité de présence élevée dans les zones de forte température moyenne annuelle), la lithologie (15.8% : probabilité de présence élevée dans la catégorie « Sédiments et volcanites ») et l'altitude (15.7% : probabilité de présence élevée dans les zones de basse altitude ; Tableau 2).



La carte de sensibilité peut également être représentée selon différentes échelles, comme par exemple à l'échelle des communes (Fig. 4 ; pour faciliter l'aide à la décision dans l'aménagement des territoires à l'échelle régionale par exemple) ou à l'échelle des bassins versants (Fig. 5).

Tableau 2. Importances relatives des variables environnementales issues du modèle prédictif. Les variables qui ont les importances relatives les plus élevées sont celles ayant une forte influence sur la probabilité de présence de l'espèce prédite par le modèle. Le sens des relations est également indiqué : ↘ représente les relations négatives entre la probabilité de présence et la variable environnementale concernée (plus la variable concernée augmente, plus la probabilité de présence de l'espèce diminue) et ↗ représente les relations positives (plus la variable concernée augmente, plus la probabilité de présence de l'espèce augmente). La variable « Lithologie » étant une variable catégorielle, le sens de la relation n'est pas indiqué

Variable	Importance relative	Sens de la relation
Température moyenne annuelle	27.1%	↗
Lithologie	15.8%	-
Altitude	15.7%	↘
Rayonnement solaire moyen annuel	9.3%	↗
Distance au plus proche espace ouvert	8.4%	↗
Distance à la plus proche autoroute	8.3%	↗
Distance au plus proche plan d'eau	5.2%	↘
Distance à la plus proche voie ferrée	3.4%	↗
Pente	2.6%	↘
Précipitation annuelle	1.5%	↗
Distance à la plus proche végétation arbustive	1.1%	↘
Distance au plus proche cours d'eau	0.9%	↘
Distance à la plus proche zone agricole	0.3%	↗
Distance à la plus proche prairie	0.3%	↘
Distance à la plus proche route secondaire	0.1%	↗
Distance à la plus proche forêt	0.1%	↘
Distance à la plus proche zone urbaine	0.0%	↘

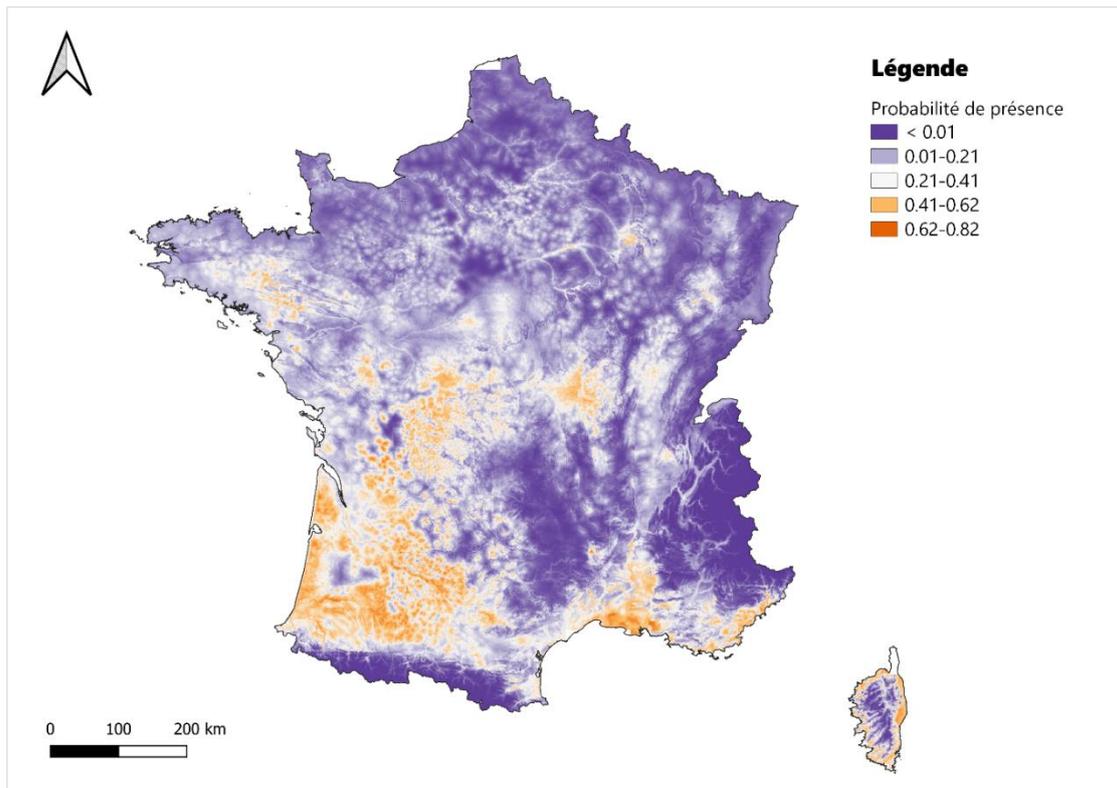


Figure 3. Carte de sensibilité nationale de la Cistude d'Europe (*Emys orbicularis*) (version de décembre 2020)

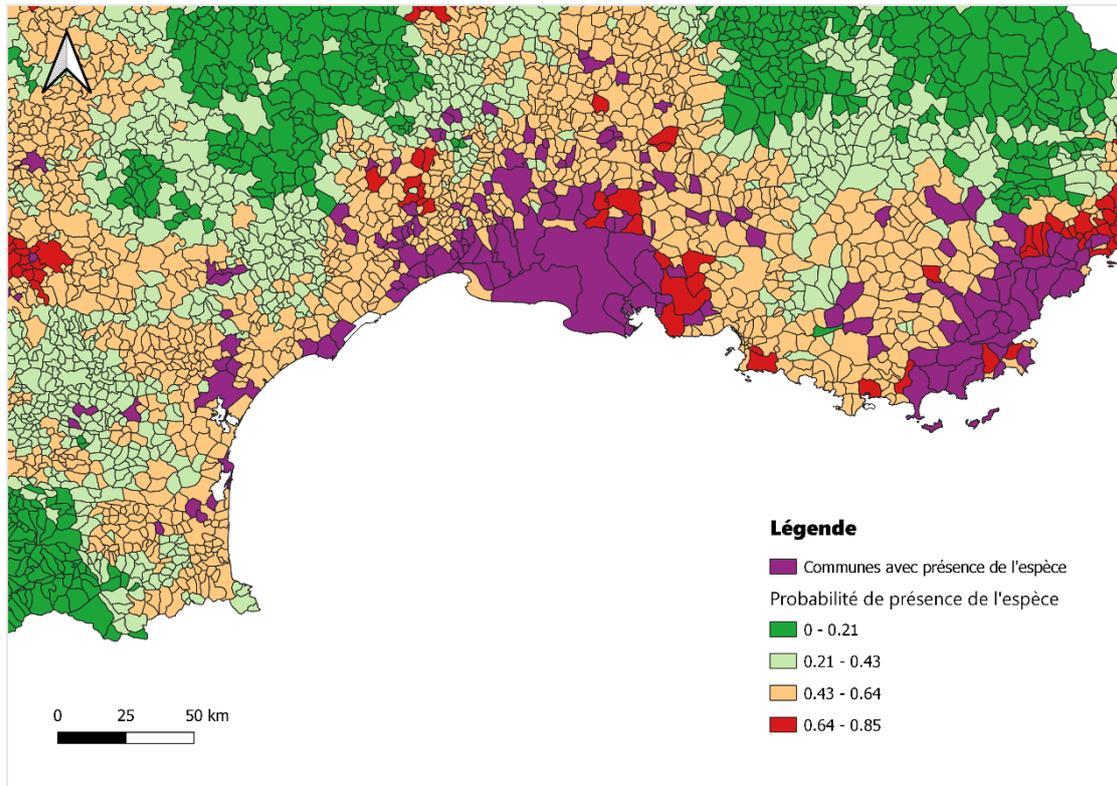


Figure 4. Carte de sensibilité nationale de la Cistude d'Europe (*Emys orbicularis*) à l'échelle des communes (zoom sur l'arc méditerranéen) (version de décembre 2020)

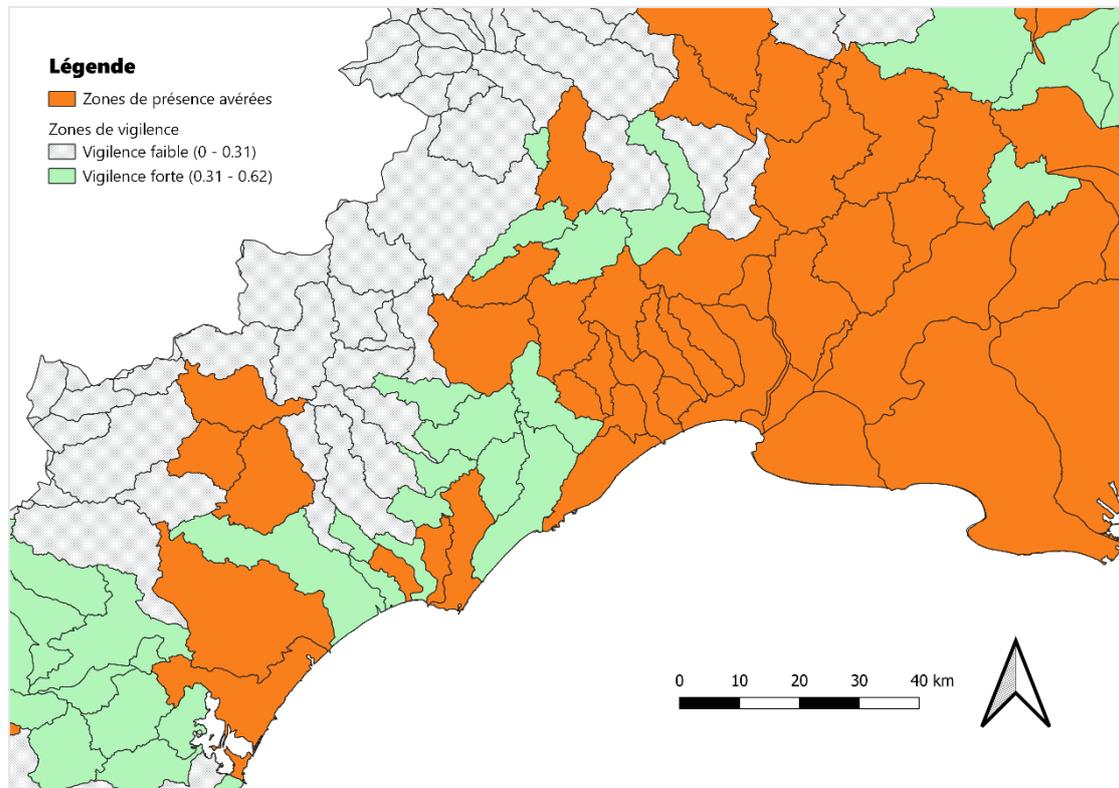


Figure 5. Carte de sensibilité nationale de la Cistude d'Europe (*Emys orbicularis*) à l'échelle des bassins versants (zoom sur l'arc méditerranéen) (version de décembre 2020)

3.4. Limites du modèle et recommandations

La réalisation d'une carte de sensibilité nationale ne remplace aucunement la connaissance des sites et des habitats à l'échelle locale. Néanmoins, la réalisation de cartes de sensibilité « manuelles » ultra-précises à très large échelle paraît impossible. Il convient donc d'utiliser cette carte de sensibilité nationale comme état des lieux des connaissances à amender, et ce, de plusieurs façons :

1. Guidé par la carte de sensibilité nationale, ainsi que de la méthodologie employée (détaillée ci-dessous), il est fortement recommandé de **réaliser des cartes de sensibilité à l'échelle régionale**. En effet, des couches d'habitats plus précises (comme l'Occupation du Sol à Grand Echelle (OCS-GE 2016), ou couches d'inventaires de zones humides par exemple) peuvent être disponibles pour certaines régions (couches non utilisables dans la modélisation nationale du fait qu'elles ne couvrent pas l'entièreté de la France). Ces couches plus précises permettent d'affiner la relation entre variables environnementales et données d'occurrence, ce qui impactera les cartes de sensibilité. Il convient donc, à l'échelle



régionale, d'utiliser des **couches d'habitats les plus précises possibles** pour la réalisation des cartes de sensibilité

2. Puis, à partir des cartes de sensibilité régionales produites, il est recommandé, dans un second temps, de se rapprocher des **experts locaux** de l'espèce considérée, afin d'amender et de valider les cartes de sensibilité produites via la modélisation prédictive.

→ La SHF se tient à disposition pour toute aide technique dans la réalisation des cartes de sensibilité régionales.

4. Bibliographie

- Araújo MB, Pearson RG & Thuiller WM (2005) Validation of species-climate impact models under climate change. *Global Change Biology* **11**: 1504–13
- Bargain A & Fabri M-C (2016) *Guide méthodologique de modélisation prédictive d'habitats profonds en méditerranée*. Département Océanographie et Dynamique des Ecosystèmes, Laboratoire Environnement Ressources Provence Azur Corse (LER-PAC). 128 pp
- Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE & Yates CJ (2011) A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions* **17**: 43-57
- Elith J, Graham CH, Anderson RP, Dudík M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BE, Manion G, Mortiz C, Nakamura M, Nakazawa Y, Overton JMO, Peterson AT, Phillips SJ, Richardson K, Scathetti-Pereira R, Schapire RE, Soberon J, Williams S, Wisz MS & Zimmermann NE (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**: 29-151
- Guisan A & Zimmermann NE (2000) Predictive habitat distribution models in ecology. *Ecological Modelling* **135**: 147-186
- Guisan A, Broennimann O, Engler R, Vust M, Yoccoz NG, Lehmann A & Zimmermann N (2005) Using niche-based models to improve the sampling of rare species. *Conservation Biology Series* **20**: 501–511
- Hernandez PA, Graham CH, Master LL & Albert DL (2006) The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* **29**: 773–785
- Kato Y (2012) *Méthodologie pour le calcul de probabilité d'implantation de l'habitat dispersé par entropie maximale*. Rapport Caribsats, programme INTERREG Caraïbe IV. 12 p.
- Luque S, Redon R, Isenmann M & Sanz T (2012) Prédiction de la distribution d'alliances de végétation des milieux ouverts d'altitude à l'aide de l'approche dite du maximum d'entropie. Programme CarHAB.
- Phillips SJ, Anderson RP & Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**: 231-259
- Pearce J & Boyce MS (2006) Modelling distribution and abundance with presence-only data. *Journal of Applied Ecology* **43**: 405–412
- Pearson RG, Raxworthy CJ, Nakamura M & Townsend Peterson A (2007) Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography* **34**: 102-117
- Préau C, Trochet A, Bertrand R & Isselin-Nondedeu F (2018) Modeling potential distributions of three European amphibian species comparing ENFA and MaxEnt. *Herpetological Conservation and Biology* **13**: 91-104



- Préau C, Lelarge K, Grandjean F, Bertrand R, Isselin-Nondedeu F & Sellier Y (2019) *Identification et modélisation des habitats d'espèces à enjeux des zones humides et évolution de leur répartition avec le changement climatique - Rapport spécial Vienne aval*. Édité par GEREPI, Vouneuil-sur-Vienne, France. 57 p.
- Préau C, Isselin-Nondedeu F, Trochet A, Bertrand R, Sansault E, Dohogne R, Jemmin J, Tinchant A, Dutertre A & Pagé G (2017) Une collaboration constructive entre associations naturalistes et scientifiques : exemple de la modélisation de la distribution des amphibiens sur un département. *Sciences Eaux & Territoires* **43** : 1-7
- Venditti V (1998) *Aspects du principe de maximum d'entropie en modélisation statistique*. Thèse de doctorat. Université de Grenoble.
- Wang, YS, Xie BY, Wan FH, Xiao QM & Dai LY (2007) The potential geographic distribution of *Radopholus similis* in China. *Agricultural Sciences in China* **6**: 1444–1449



5. Guide technique pour la réalisation des cartes de sensibilité (méthode MaxEnt)

5.1. Données d'occurrence

5.1.1. Récolte des données

Dans un premier temps, il convient de s'assurer que le séparateur de décimales soit en version anglo-saxonne (séparés par des points). *Sous Windows 10 : Paramètres > Heures et langues > Région > Paramètres associés (Date, heure & paramètres régionaux supplémentaires) > Région > Modifier les formats de date, d'heure et de nombre > Formats > Paramètres supplémentaires > Symbole décimal « . » au lieu de « , ».*

Pour la réalisation des cartes de sensibilité, et ce, sur n'importe quel territoire, il est impératif de récolter la **majorité des données d'occurrence** (points géographiques d'observations) de l'espèce sur l'échelle spatiale considérée (locale, régionale, nationale...). En effet, la robustesse du modèle prédictif dépendra fortement des données d'occurrence qui y auront été saisies, puisque ces dernières seront mises en relation avec les variables environnementales. Par conséquent, l'omission d'une partie des points d'observation de l'espèce (sur un département par exemple) pourrait biaiser les « vraies » conditions environnementales favorables de l'espèce estimées par le modèle prédictif. Ainsi, la carte de sensibilité produite risquera d'être peu réaliste (faible valeur d'AUC et zones de vigilance aberrantes).

Les données à compiler doivent impérativement être des **données pointées** (coordonnées X et Y). Toutes les autres représentations (données maillées ou à l'échelle de la commune) sont à supprimer de la base de données pour éviter tout biais dans l'analyse.

La compilation des données d'occurrence peut se faire en combinant les données d'inventaires de multiples structures locales (associations) et/ou via des demandes d'export de synthèses régionales (SINP) et/ou nationale (INPN, <https://inpn.mnhn.fr/espece/extraction-sinp/preamble>).

5.1.2. Formatage des données

Le fichier des données d'occurrence ne doit contenir que des données de présence avérées (aucune données d'absence). Par conséquent, les utilisateurs s'engagent à vérifier/contrôler toutes les données d'occurrence (voire supprimer certaines anciennes données non revalidées depuis par exemple) compilées avant de les implémenter dans le modèle prédictif.

Une fois réceptionnées, les données doivent être compilées dans un seul fichier (Excel) avec 2 colonnes (une ligne représentant une observation, une colonne pour la coordonnée X et une colonne pour la coordonnée Y). Le fichier doit ensuite être enregistré sous **format .txt** (sous Excel > Enregistrer sous > « nomdufichier » .txt).

X	Y
330949.03	6260457.09
335610.57	6269909.55

Format du fichier contenant les données d'occurrence (enregistré en .txt)

→ Il n'y a pas de système de projection obligatoire dans le formatage des fichiers, les coordonnées peuvent donc être indiquées selon le système de projection WGS84, ou Lambert 93 par exemple. Il faut néanmoins veiller à ce que le système de projection entre le fichier des points d'occurrence et les fichiers des variables environnementales soit identique. Les coordonnées doivent cependant être indiquées sous **format décimal** (et non en degrés, minutes, secondes).

5.2. Variables environnementales

Les variables environnementales choisies sont des variables susceptibles d'influencer (positivement ou négativement) la distribution de l'espèce considérée. Ces variables environnementales vont décrire quantitativement ou qualitativement les conditions environnementales à chaque point géoréférencé de présence de l'espèce.

5.2.1. Téléchargement des couches habitats

Plusieurs couches d'habitats sont disponibles en téléchargement libre :



- ➔ Téléchargement du Corine Land Cover (CLC) en format vecteur (.shp) à l'échelle de la France : <https://www.data.gouv.fr/fr/datasets/corine-land-cover-occupation-des-sols-en-france/>
- ➔ Téléchargement de la couche Occupation du Sol à Grand Échelle (OCS-GE 2016) en format vecteur (.shp) pour certains départements ou certaines régions (uniquement de France) : <https://geoservices.ign.fr/documentation/diffusion/telechargement-donnees-libres.html#ocs-ge>
- ➔ Téléchargement de la couche OSO, produite par le CESBIO, en format raster (.tiff) à l'échelle de la France : <http://osr-cesbio.ups-tlse.fr/~oso/>
- ➔ Téléchargement du Modèle Numérique de Terrain (MNT) en format raster (.tif) à l'échelle de la France : <https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1.1?tab=mapview>
- ➔ Téléchargement de données climatiques mondiales en format raster (.tif) : températures minimales, moyennes et maximales, précipitations et rayonnement solaire entre 1970 et 2000 (12 fichiers par variables, un par mois) : <https://www.worldclim.org/>
- ➔ Téléchargement du réseau routier (Route 500®) en format vecteur (.shp) sur le site de l'IGN à l'échelle de la France : <https://geoservices.ign.fr/documentation/diffusion/telechargement-donnees-libres.html>
- ➔ Téléchargement du réseau ferré (SNCF) en format vecteur (.shp) à l'échelle de la France : <https://www.data.gouv.fr/fr/datasets/fichier-de-formes-des-lignes-du-reseau-ferre-national/>
- ➔ Téléchargement de la BD Carthage (cours d'eau et point d'eau) en format vecteur (.shp) à l'échelle de la France : <https://www.data.gouv.fr/fr/datasets/plans-deau-metropole-2016-bd-carthage/>

5.2.2. Création des rasters « habitats »

a. Création des cartes d'habitats en format raster (.tif) [QGIS]

L'objectif est de réunir un lot de plusieurs couches d'habitats susceptibles d'avoir une influence (positive ou négative) sur la présence de l'espèce cible (par exemple chez les amphibiens, la présence de zones humides est importante pour la reproduction, cette variable a donc de fortes chances de conditionner la présence des espèces). Le choix des couches habitats dépendra donc de l'espèce considérée.

Les couches habitats (en format raster et/ou vecteur) ainsi que le fichier contenant les coordonnées des observations de l'espèce considérée sont obligatoirement à mettre dans le même système de projection. Ici, les analyses sont réalisées dans le système de projection Lambert 93 (EPSG : 2154).

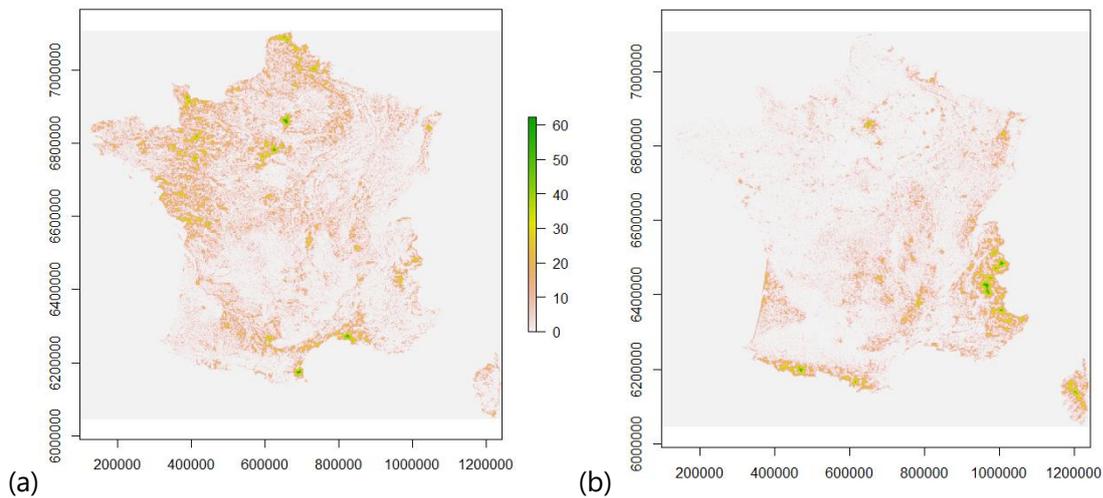


Pour les variables de présence de certains habitats (comme la présence d'une zone forestière, ou d'un cours d'eau), un calcul supplémentaire a été effectué au moment de la création des fichiers raster : dans chaque pixel des rasters créés, plutôt que d'associer uniquement deux valeurs (0 pour absence de l'habitat ou 1 pour présence de l'habitat), la distance au plus proche habitat est calculée.

POUR LA COUCHE CLC (FORMAT VECTEUR) :

- Ouvrir le fichier dans QGIS
- **Vérifier et, si besoin, changer le système de projection** : clic droit sur la couche → Exporter → Sauvegarder les entités sous → indiquer le nom de la nouvelle couche (par exemple « CLC_L93.shp ») et indiquer le SCR « EPSG : 2154 - Lambert-93 »
- **Sélection d'habitats** : sélectionner une catégorie d'entités suivant le code CLC (par exemple, pour une carte d'habitats incluant les zones urbaines, sélectionner dans la table attributaire toutes les entités ayant un code commençant par 100 dans la colonne correspondante)
- **Enregistrer les entités sélectionnées dans une nouvelle couche** : créer une nouvelle couche avec les entités sélectionnées (par exemple : clic droit sur la couche → Exporter → Sauvegarder les entités sélectionnées sous → indiquer le nom de la nouvelle couche (par exemple « CLC_L93_zones_urbaines.shp ») et indiquer le SCR « EPSG : 2154 - Lambert-93 »
- **Ajouter une colonne** : créer une nouvelle colonne dans la table attributaire de la nouvelle couche (Activer le mode édition → Ajouter un champ « CODE » en nombre entier) et attribuer une valeur de 10 à toutes les entités (en haut de la table attributaire : abc ID = 10)
- **Fusionner la couche** : fusionner la nouvelle couche avec celle du contour administratif de la France (qui doit également être dans le système de projection Lambert 93) : Vecteur → Outils de gestion de données → Fusionner des couches vecteur (attention, la couche du contour administratif de la France doit être positionnée en dessous de la couche de la nouvelle couche dans la fenêtre de fusion : le fichier temporaire créé doit faire apparaître les éléments d'habitats sélectionnés avec le contour administratif de la France)
- **Transformer le fichier vecteur en format raster** (la résolution choisie ici est une taille de pixel de 200 x 200 m) : Raster → Conversion → Rastérisation [champ à utiliser : CODE, valeur fixe : 1, Unité : Unités géoréférencées, Largeur/résolution horizontale : 200, Largeur/résolution verticale : 200, Emprise du résultat : fichier vecteur du contour administratif de la France, Affecter une valeur nulle : 1].
- **Attribuer une valeur dans chaque pixel** : calculer la distance euclidienne à l'entité la plus proche et attribuer cette valeur dans chaque pixel du raster : Raster → Analyse → Proximité (fichier temporaire créé)

- **Enregistrer le fichier** : enregistrer le raster : clic droit → Exporter → Enregistrer sous (format .tif, « EPSG : 2154 - Lambert-93 »).
- **Répéter la même opération** pour chaque catégorie d'habitats du CLC désirée



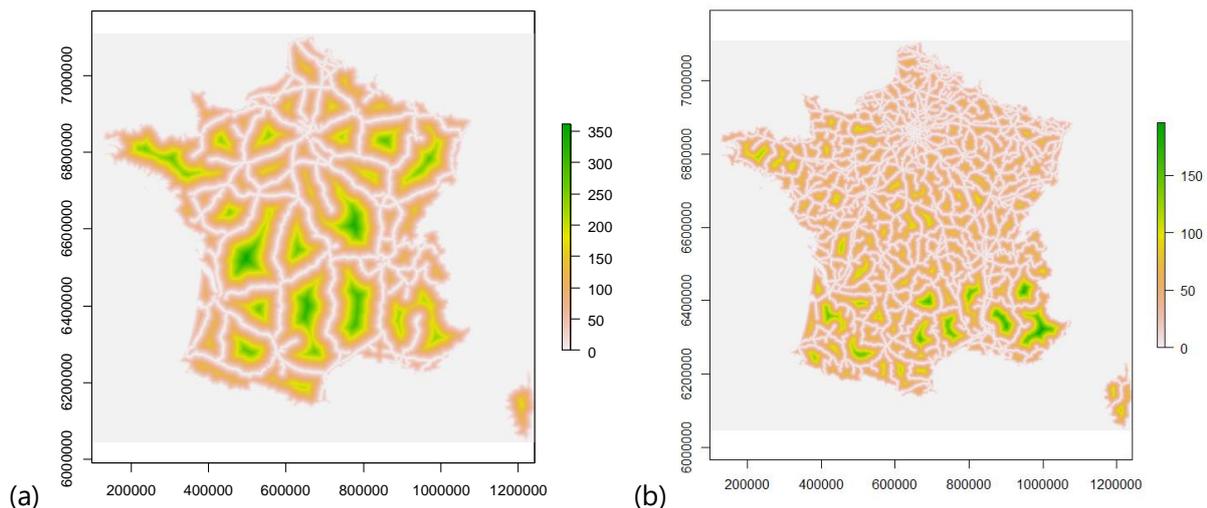
Exemples des fichiers rasters représentant (a) la proximité aux forêts et (b) la proximité aux zones agricoles. Les zones vertes correspondent à des localités éloignées des deux types d'habitats (données issues du CLC)

POUR LES COUCHES LINÉAIRES (FORMAT VECTEUR) COMME LES ROUTES :

- Ouvrir le fichier dans QGIS
- **Vérifier et, si besoin, changer le système de projection** : clic droit sur la couche → Exporter → Sauvegarder les entités sous → indiquer le nom de la nouvelle couche (par exemple « Routes_L93.shp ») et indiquer le SCR « EPSG : 2154 - Lambert-93 »
- **Transformer le fichier vecteur lignes en polygones** : Vecteur → Outils de géométrie → De Lignes à Polygones (fichier temporaire créé) ou faire une **Mise en Tampon** : Vecteur → Outils de géotraitement → Tampon (fichier temporaire créé)
- **Ajouter une colonne** : créer une nouvelle colonne dans la table attributaire de la nouvelle couche (Activer le mode édition → Ajouter un champ « CODE » en nombre entier) et attribuer une valeur de 10 à toutes les entités (en haut de la table attributaire : abc ID = 10)
- **Fusionner la couche** : fusionner la nouvelle couche avec celle du contour administratif de la France (qui doit également être dans le système de projection Lambert 93) : Vecteur → Outils de gestion de données → Fusionner des couches vecteur (attention, la couche du contour administratif de la France doit être positionnée en dessous de la couche de la

nouvelle couche dans la fenêtre de fusion : le fichier temporaire créé doit faire apparaître les éléments d'habitats sélectionnés avec le contour administratif de la France)

- **Transformer le fichier vecteur en format raster** (la résolution choisie ici est une taille de pixel de 200 x 200 m) : Raster → Conversion → Rastérisation [champ à utiliser : CODE, valeur fixe : 1, Unité : Unités géoréférencées, Largeur/résolution horizontale : 200, Largeur/résolution verticale : 200, Emprise du résultat : fichier vecteur du contour administratif de la France, Affecter une valeur nulle : 1].
- **Attribuer une valeur dans chaque pixel** : calculer la distance euclidienne à l'entité la plus proche et attribuer cette valeur dans chaque pixel du raster : Raster → Analyse → Proximité (fichier temporaire créé)
- **Enregistrer le fichier** : enregistrer le raster : clic droit → Exporter → Enregistrer sous (format .tif, « EPSG : 2154 - Lambert-93 »).

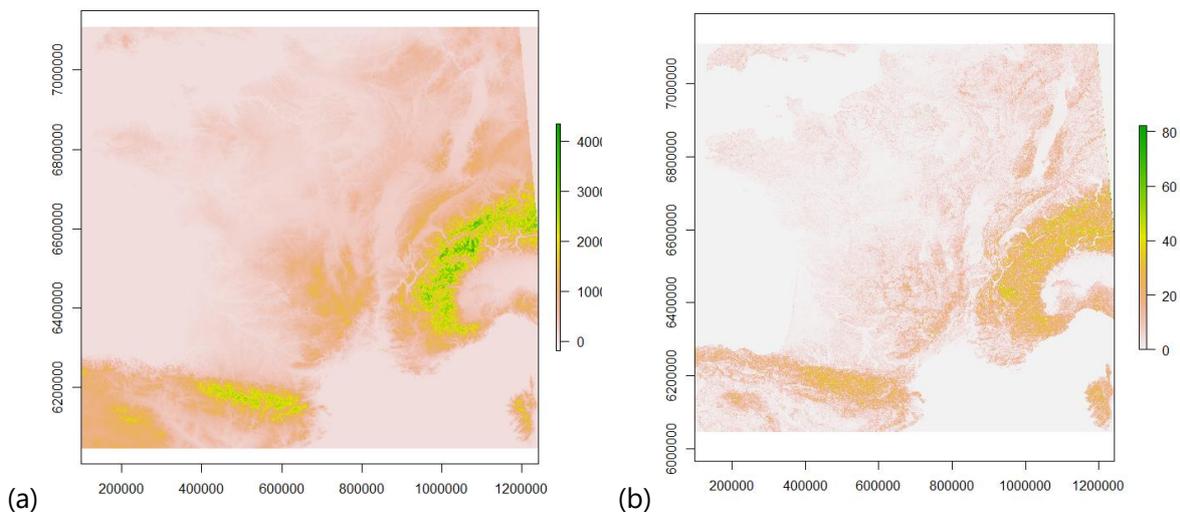


Exemples des fichiers rasters représentant (a) la proximité des autoroutes et (b) la proximité des voies ferrées. Les zones vertes correspondent à des localités éloignées des deux types d'habitats

POUR LA COUCHE MNT (FORMAT RASTER NON CHEVAUCHANT) :

- Ouvrir le fichier dans QGIS
- **Créer un seul et unique fichier raster** : fusionner les couches raster du MNT (Raster → Divers → Fusionner)
- **Calculer la Pente** : Raster → Analyse → Pente (fichier temporaire créé)
- **Calculer la Rugosité** : Raster → Analyse → Rugosité du terrain (fichier temporaire créé)

- **Enregistrer le fichier** (un fichier à enregistrer à partir du raster unique de MNT pour obtenir l'altitude, puis un fichier à enregistrer à partir du fichier temporaire pour la pente, et un autre fichier à enregistrer pour la rugosité) : enregistrer en raster .tif (clic droit sur la couche CLC → Exporter → Sauvegarder les entités sous → indiquer le nom de la nouvelle couche (par exemple, « Altitude.tif »), indiquer le SCR « EPSG : 2154 - Lambert-93 » et indiquer l'emprise désirée (fichier vecteur du contour administratif de la France), qui doit être la même pour tous les fichiers



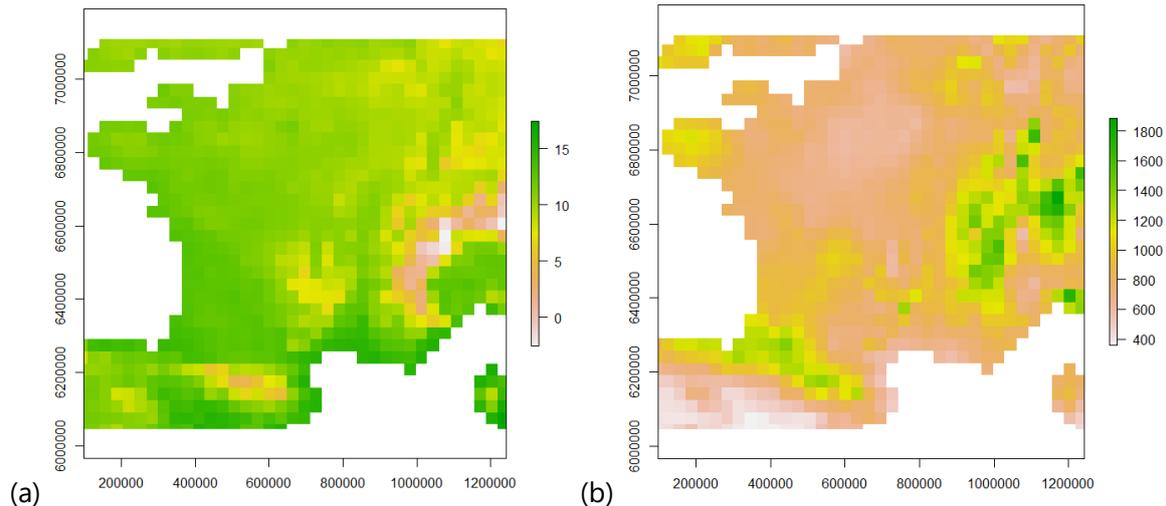
Exemples des fichiers rasters représentant (a) l'altitude et (b) la pente. Les zones vertes correspondent à des localités ayant une altitude ou une pente élevée

POUR LES COUCHES CLIMATIQUES (FORMAT RASTER CHEVAUCHANT) :

- Ouvrir le fichier dans QGIS
- **Vérifier et, si besoin, changer la projection** : mettre tous les fichiers raster un par un dans le même système de projection (clic droit sur la couche → Exporter → Sauvegarder les entités sous → indiquer le nom de la nouvelle couche (par exemple, « Tmin_93 ») et indiquer le SCR « EPSG : 2154 - Lambert-93 »)
- **Couper les rasters selon l'emprise souhaitée** : enregistrer les fichiers raster un par un selon l'emprise de la couche souhaitée (fichier vecteur du contour administratif de la France) avec une résolution de 200 m x 200 m : clic droit sur la couche → Exporter → Sauvegarder les entités sous → indiquer le nom de la nouvelle couche (par exemple « Tmin_93_France »), indiquer l'emprise de la couche (fichier vecteur du contour administratif de la France) et la résolution (200 x 200)



- **Extraire les valeurs des rasters** : Extraire les valeurs min, max et/ou moyenne des douze rasters chevauchants : Traitement → Boite à outils → Statistics for raster (en cochant MIN, MAX ou MEAN)
- **Enregistrer le fichier** : enregistrer le fichier raster en format .tif (clic droit → Exporter → Enregistrer sous (format .tif, « EPSG : 2154 - Lambert-93 »).



Exemples des fichiers rasters représentant (a) la température moyenne annuelle et (b) les précipitations annuelles. Les zones vertes correspondent à des localités ayant une température moyenne annuelle ou une valeur de précipitation annuelle élevée

b. Transformation des couches habitats .tif en .grd [R]

→ Dans R, il faut éviter les espaces (dans les noms des fichiers, dossiers, etc.)

→ Les indications en oranges doivent être modifiées en fonction de vos noms de fichiers/dossiers

ouvrir une session sous R

installer la librairie "raster" (à faire une seule fois)

```
install.packages("raster")
```

charger la librairie "raster" (à faire à chaque fois)

```
library(raster)
```

indiquer le chemin vers le dossier contenant toutes les couches habitats .tif

```
setwd("C:/Users/.../Cartes_sensibilité/Cistude/habitats")
```

charger une première couche habitat .tif

```
r1 <- raster("Plan_deau_dist.tif")
```

vérifier la projection

```
r1
```

si besoin, changer la projection (toutes les couches habitats, ainsi que le fichier contenant les points d'observation de l'espèce) doivent être dans la même projection



```
proj4string(r1) <- CRS("+init=epsg:2154") # transformation en Lambert 93
# enregistrer le fichier raster en .grd
writeRaster(r1, "Plan_deau_dist.grd") # attention ici à renommer correctement le fichier .grd créé
(pour éviter d'écraser les autres fichiers habitats)
# répéter l'opération pour tous les fichiers rasters créés précédemment
# mettre tous les fichiers .grd dans le même dossier
```

5.3. Modélisation prédictive d'habitats sous R

5.3.1. MaxEnt

→ Dans R, il faut éviter les espaces (dans les noms des fichiers, dossiers, etc.)
→ Les indications en oranges doivent être modifiées en fonction de vos noms de fichiers/dossiers

```
# ouvrir une session sous R
# installer les librairies suivantes (à faire une seule fois)
install.packages("raster")
install.packages("devtools")
install.packages("rmaxent")
install.packages("raster")
install.packages("dismo")
install.packages("rJava")
install.packages("rasterVis")
install.packages("viridis")
# charger les librairies suivantes (à faire à chaque fois)
library(devtools)
install_github('johnbaums/rmaxent')
library(rmaxent)
get_maxent('3.3.3k')
library(raster)
library(dismo)
library(rJava)
library(raster)
library(rasterVis)
library(viridis)
# indiquer le chemin vers le dossier contenant le fichier .txt des points d'observation
setwd(dir = "C:/Users/.../Cartes_sensibilité/Cistude")
#----- charger le fichier .txt avec tous les points d'observation (avec 2 colonnes :
coordonnées X et Y en Lambert 93)
occ <- read.table("obs_cistude.txt", h=T) # le nombre des colonnes ne doit pas contenir d'espaces et
les valeurs des coordonnées doivent être en décimale avec un point et non une virgule
```



```
#----- indiquer le chemin vers le dossier contenant toutes les couches habitats .tif
pred_files <- list.files("C:/Users/.../Cartes_sensibilite/Cistude/habitats", patt = "\\*.grd", full.names
=TRUE)
pred_files
predictors <- stack(pred_files)
#----- modèle MaxEnt (après "factors=" indiquer les couches habitats catégorielles, comme
par exemple les couches géologiques (qui indiquent les zones calcaires, volcaniques, etc.). Si aucune
couche n'est catégorielle, alors supprimer "factors='Lithologie',")
me <- maxent(predictors, occ, factors='Lithologie', args=c('hinge=false', 'threshold=false')) # après
"factors=", indiquer le nom des variables catégorielle (ici Lithologie)
me # ouvre une page internet avec les résultats de la modélisation : en bas de page sont indiquées
les importances relatives de chaque variable environnementale
# représentation graphique de l'importance relative de chaque couche d'habitats
plot(me)
# représentation graphique de chaque variable (montre le sens des relations (positives ou négatives)
entre la probabilité de présence de l'espèce et chaque variable environnementale)
response(me)
# augmenter la taille de la capacité de la mémoire du logiciel
memory.size(max =20000)
#----- création du modèle prédictif
prediction <- project(me, predictors)
# représentation cartographique du modèle prédictif seul
levelplot(prediction$prediction_logistic, margin=FALSE, col.regions=viridis, at=seq(0, 1, len=100))
# représentation cartographique du modèle prédictif avec les points d'occurrence
levelplot(prediction$prediction_logistic, margin=FALSE, col.regions=viridis, at=seq(0, 1, len=100))+
layer(sp.points(SpatialPoints(occ), pch=20, col=2))
#----- enregistrer le modèle prédictif (en format raster)
writeRaster(prediction$prediction_logistic, filename=file.path("C:/Users/.../Cartes_sensibilité/
Cistude/sorties", "sortie_maxent.tif"))
```

5.3.2. Évaluation de la robustesse du modèle prédictif

→ Les indications en oranges doivent être modifiées en fonction de vos noms de fichiers/dossiers

```
# à réaliser après avoir lancé le modèle prédictif ci-dessus
# installer les bibliothèques suivantes (à faire une seule fois)
install.packages('biomod2')
install.packages('raster')
install.packages('dismo')
install.packages('reshape')
# charger les bibliothèques suivantes (à faire à chaque fois)
library(biomod2)
library(raster)
```



```

library(dismo)
library(reshape)
#----- charger le jeu de données des points d'observations
setwd(dir = "C:/Users/.../Cartes_sensibilite/Cistude")
DataSpecies=read.table("obs_cistude_2.txt", h=T)
myRespName <- 'Emys_orbicularis'
myResp <- as.numeric(DataSpecies[,myRespName])
head(myResp)
myRespXY <- DataSpecies[,c("X","Y")]
head(myRespXY)
#----- charger les fichiers habitats
setwd(dir = "C:/Users/.../Cartes_sensibilite/Cistude/habitats")
habitats_liste=list.files("C:/Users/.../Cartes_sensibilite/Cistude/habitats", patt = "\\.*grd")
myExpl=stack(list.files("C:/Users/.../Cartes_sensibilite/Cistude/habitats", patt = "\\.*grd"))
#----- Setting up data file for Biomod2
myBiomodData <- BIOMOD_FormatingData(resp.var = myResp,
                                   expl.var = myExpl,
                                   resp.xy = myRespXY,
                                   resp.name = myRespName,
                                   PA.nb.rep = 1, # Est-ce qu'il faut des pseudo-absences ? (O=non, 1=oui)
                                   PA.nb.absences = 1000, # nombre de pseudo-absences
                                   PA.strategy = 'random')

myBiomodData
# 11356 presence data + 699 pseudo-absences + 17 couches habitats
#----- Computing the models and running Maxent
myBiomodOption <- BIOMOD_ModelingOptions()
myBiomodModelOut <- BIOMOD_Modeling(myBiomodData,
                                   models = c('MAXENT.Phillips'),
                                   models.options = myBiomodOption,
                                   NbRunEval=10,
                                   DataSplit=80,
                                   Prevalence=0.5,
                                   VarImport=3,
                                   models.eval.meth = c('TSS','ROC'),
                                   SaveObj = TRUE,
                                   rescal.all.models = TRUE,
                                   do.full.models = FALSE,
                                   modeling.id = paste(myRespName,"FirstModeling",sep=""))
# ----- Evaluation du modèle
myMaxentModelEval <- get_evaluations(myBiomodModelOut, as.data.frame = T)
myMaxentModelEval
ROC=myMaxentModelEval[myMaxentModelEval$Eval.metric=="ROC",]
mean(ROC$Testing.data)

```



AUC = 0.8702

5.4. Fichier de sortie raster [QGIS]

- **Ouvrir** le fichier de sortie Maxent sous QGIS
- Traitement → Boîte à outils → Zonal Statistics (de nouvelles colonnes apparaissent dans le fichier des communes avec les valeurs min, max, moyennes des valeurs du raster)
- Clic droit sur la couche → Propriétés → Symbologie → Gradué (sur valeurs moyennes) → Appliquer. Les communes en rouge sont celles avec la plus forte probabilité de présence de l'espèce considérée