

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/330970900>

# Modélisation spatiale de la probabilité de présence relative du Lézard ocellé dans les régions Provence-Alpes-Côte d'Azur et ex-Languedoc-Roussillon

Technical Report · January 2018

DOI: 10.13140/RG.2.2.18107.05923

CITATIONS

0

READS

201

5 authors, including:



**Guillaume Astruc**

Centre d'Ecologie Fonctionnelle et Evolutive

27 PUBLICATIONS 188 CITATIONS

[SEE PROFILE](#)



**Michaël Guillon**

BIOTOPE Bureau d'études environnement et milieux naturels faune flore

27 PUBLICATIONS 557 CITATIONS

[SEE PROFILE](#)



**Olivier Lourdaïs**

French National Centre for Scientific Research

126 PUBLICATIONS 2,900 CITATIONS

[SEE PROFILE](#)



**Marc-Antoine Marchand**

Conservatoire d'espaces naturels de Provence-Alpes-Côte d'Azur

24 PUBLICATIONS 76 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Adult lizard activity pattern by camera traps [View project](#)



Long-Term Evaluation of the Success of a Reintroduction Program of the European Pond Turtle [View project](#)

# Modélisation spatiale de la probabilité de présence relative du Lézard ocellé dans les régions Provence-Alpes-Côte d'Azur et ex-Languedoc-Roussillon

## Déclinaison Inter-Régional PACA et LR du PNA Lézard ocellé



École Pratique  
des Hautes Études

**Ecole Pratique des Hautes Etudes**  
46, rue de Lille - 75007 Paris



**Centre d'Ecologie Fonctionnelle et Evolutive UMR 5175**  
1919 Route de Mende - 34293 Montpellier



Centre d'Études  
Biologiques de  
**Chizé**

**Centre d'Etude Biologique de Chizé UMR 7372**  
405 route de la Canauderie - 79360 Villiers en Bois

# Modélisation spatiale de la probabilité de présence relative du Lézard ocellé dans les régions Provence-Alpes-Côte d'Azur et ex-Languedoc-Roussillon

## Déclinaison Inter-Régional PACA et LR du PNA Lézard ocellé

### Document réalisé par :



Conservatoire d'espaces naturels de Provence-Alpes-Côte d'Azur.



Ecole Pratique des Hautes Etudes – UMR5175 Centre d'Ecologie Fonctionnelle et Evolutive.



Centre d'Etude Biologique de Chizé – UMR 7372

### Avec le soutien financier de :



Direction Régionale de l'Environnement, de l'Aménagement et du Logement de Provence - Alpes - Côte d'Azur.



Direction Régionale de l'Environnement, de l'Aménagement et du Logement d'Occitanie.

### Rédaction :

Guillaume Astruc - EPHE – UMR5175 CEFE  
Michaël Guillon - CEBC – UMR 7372  
Olivier Lourdais - CEBC – UMR 7372  
Marc-Antoine Marchand - CEN PACA  
Aurélien Besnard - EPHE – UMR5175 CEFE

**Date de réalisation :** 2018

### Crédits photographiques :

1<sup>ère</sup> de couverture : Couple de Lézard ocellé © J. Celse (CEN PACA)  
Pour le reste des illustrations, l'auteur est mentionné dans la légende

### Citation recommandée :

Astruc G., Guillon M., Lourdais O., Marchand M-A. & Besnard A. 2018. Modélisation spatiale de la probabilité de présence relative du Lézard ocellé dans les régions Provence-Alpes-Côtes d'Azur et ex-Languedoc-Roussillon - Déclinaison Inter-Régionale PACA et LR du PNA Lézard ocellé. Conservatoire d'espaces naturels de Provence-Alpes-Côte d'Azur. Sisteron. 27p.



## Sommaire

<b>CONTEXTE .....</b>	<b>2</b>
<b>OBJECTIFS .....</b>	<b>3</b>
<b>METHODES.....</b>	<b>4</b>
Principe des modèles de distribution :.....	4
Données nécessaires :.....	4
Traitement des données : .....	6
Les variables retenues pour l'analyse : .....	6
Modèles utilisés :.....	7
Prise en compte de l'effort d'échantillonnage .....	7
<b>RESULTATS ET DISCUSSIONS.....</b>	<b>8</b>
Contribution des variables au modèle de distribution potentielle.....	13
Courbes de réponses des variables utilisées pour modéliser la distribution du Lézard ocellé .....	15
Tests de validation du modèle de distribution .....	17
Confrontation des données d'observation au modèle de distribution .....	19
<b>INTERPRETATION ET PRECAUTIONS D'UTILISATION DE L'OUTIL .....</b>	<b>21</b>
<b>PERSPECTIVES .....</b>	<b>22</b>
<b>BIBLIOGRAPHIE .....</b>	<b>23</b>
<b>ANNEXES.....</b>	<b>24</b>

## Table des figures

Figure 1 : Nombre de projets faisant l'objet d'arrêtés de dérogation concernant le Lézard ocellé par année dans l'ex-région Languedoc-Roussillon (source : ex-DREAL Languedoc-Roussillon, 29/09/2017). .....	3
Figure 2 : Recherche ciblée sur le Lézard ocellé dans un secteur d'altitude dans le département des Alpes-Maritimes. ....	3
Figure 3 : Localisation des observations de Lézard ocellé de 2007 à 2016 en régions PACA et ex-LR centralisées dans le cadre du PIRA LO et d'après la base de données MALPOLON gérée par l'EPHE. Fond de carte : BDALTI@IGN. ....	5
Figure 4 : Probabilité de présence relative du Lézard ocellé avec prise en compte de l'effort d'échantillonnage. De 0 à 1 : probabilité de présence relative faible à forte. ....	9
Figure 5 : Probabilité de présence relative du Lézard ocellé sans prise en compte de l'effort d'échantillonnage. De 0 à 1 : probabilité de présence relative faible à forte. ....	10
Figure 6 : Coefficient de variation de la probabilité de présence relative du Lézard ocellé sans prise en compte de l'effort d'échantillonnage. Plus la valeur du coefficient de variation est élevée, plus l'incertitude est forte (pour plus de précisions sur la méthode de calcul des coefficients de variations : <a href="https://cran.r-project.org/web/packages/biomod2/biomod2.pdf">https://cran.r-project.org/web/packages/biomod2/biomod2.pdf</a> ). ....	11
Figure 7 : Coefficient de variation de la probabilité de présence relative du Lézard ocellé avec prise en compte de l'effort d'échantillonnage. Plus la valeur du coefficient de variation est élevée, plus l'incertitude est forte (pour plus de précisions sur la méthode de calcul des coefficients de variations : <a href="https://cran.r-project.org/web/packages/biomod2/biomod2.pdf">https://cran.r-project.org/web/packages/biomod2/biomod2.pdf</a> ). ....	12
Figure 8 : Contribution moyenne de chaque variable testée au modèle de distribution final. Les boîtes d'erreurs représentent l'intervalle de confiance à 95% des contributions obtenues sur les 100 modèles. ....	13
Figure 9 : A : Variable représentant les degrés / jours cumulés. De 0 à 40 000 : température basse à haute. B : Probabilité de présence relative du Lézard ocellé avec prise en compte de l'effort d'échantillonnage. De 0 à 1 : probabilité de présence relative faible à forte. ....	14
Figure 10 : Courbes de réponses des 14 variables retenues pour prédire la probabilité de présence relative du Lézard ocellé. Pour chaque variable sont représentées les courbes des 100 GAMs générés par les tirages aléatoires différents. Sur chaque graphique, verticalement se lit la probabilité de présence relative du Lézard ocellé (de 0 à 1) et horizontalement la valeur de la variable (de gauche à droite : valeurs faibles à fortes). Encadrés orange : variables ayant le plus de poids dans la distribution du modèle. ....	16
Figure 11 : Moyenne et intervalle de confiance à 95% des valeurs des tests de validation des modèles de distribution sur les 100 modèles ajustés. ....	18
Figure 12 : Superposition du modèle de distribution potentielle et des données d'observations au sud de la Lozère et au nord du Gard. Points noirs : observation de Lézard ocellé ; de 0 à 1 : probabilité de présence relative faible à forte. ....	19
Figure 13 : Superposition du modèle de distribution potentielle et des données d'observations sur le secteur de la plaine des Maures jusqu'au Préalpes d'Azur. Points noirs : observation de Lézard ocellé ; de 0 à 1 : probabilité de présence relative faible à forte. ....	20



## Contexte

Le Plan national d'action (PNA) en faveur du Lézard ocellé (Doré & Thirion, 2012) a été décliné au niveau des régions Provence-Alpes-Côte d'Azur et ex-Languedoc-Roussillon. Un Plan Inter Régional d'Actions (PIRA) spécifique à ces deux régions a été élaboré et mis en œuvre de 2013 à 2017 (Legouez & Marchand, 2013).

Les résultats présentés ci-après correspondent à la mise en œuvre de l'action 2 du PIRA : Caractériser les habitats du Lézard ocellé et modéliser une répartition potentielle. La réalisation de cette action contribue à la réalisation des actions 4 : Actualiser la répartition du Lézard ocellé et 8 : Etablir une méthodologie pour éviter, réduire et compenser.

Ce travail est le résultat d'une collaboration entre le Conservatoire d'Espaces Naturels de PACA, l'équipe Ecologie et Biogéographie des Vertébrés de l'École Pratique des Hautes Etudes et du Centre d'Ecologie Fonctionnelle et Evolutive de Montpellier et le Centre d'Etude Biologique de Chizé.

Cette étude a été initiée en accord avec les DREALs des régions PACA et ex-Languedoc-Roussillon ainsi qu'avec les membres du comité de pilotage du PIRA Lézard ocellé qui s'est tenu le 25 février 2014 à la station biologique de la Tour du Valat. Elle a été portée à connaissance lors du comité de pilotage du PNA tenu le 24 mars 2016.

**Le travail présenté dans ce document est destiné à être poursuivi et à être amélioré au cours du temps et de l'accroissement des données disponibles. Ainsi, tout lecteur souhaitant faire part de ses commentaires-critiques y est vivement invité en écrivant à : [marcnatoine.marchand@cen-paca.org](mailto:marcnatoine.marchand@cen-paca.org).**

## Objectifs

**1/** Fournir aux politiques publiques et autres acteurs de la conservation un outil permettant une meilleure prise en compte du Lézard ocellé dans les projets d'aménagements du territoire en croissance (cf. Figure 1).

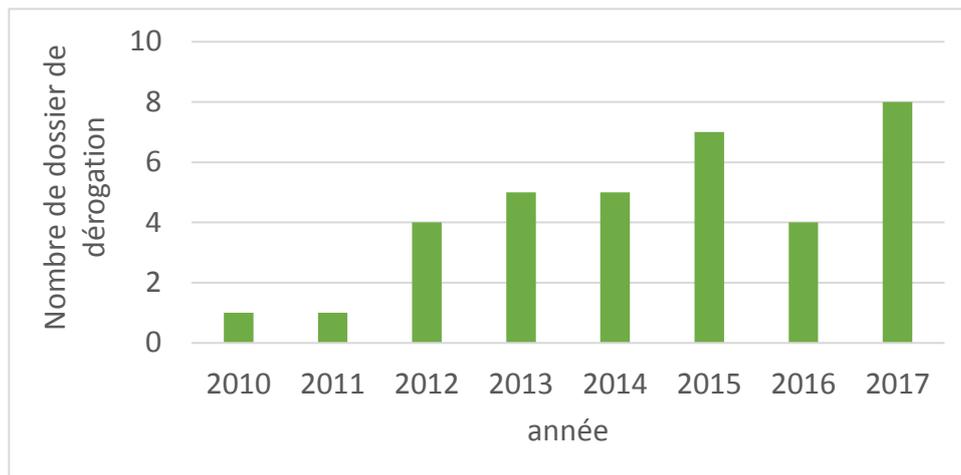


Figure 1 : Nombre de projets faisant l'objet d'arrêtés de dérogation concernant le Lézard ocellé par année dans l'ex-région Languedoc-Roussillon (source : ex-DREAL Languedoc-Roussillon, 29/09/2017).

**2/** Fournir une aide aux naturalistes et gestionnaires d'espaces naturels pour la réalisation de prospections ciblées (cf. Figure 2) afin d'améliorer les connaissances sur la distribution de l'espèce.



Figure 2 : Recherche ciblée sur le Lézard ocellé dans un secteur d'altitude dans le département des Alpes-Maritimes.

**3/** Améliorer les connaissances sur l'écologie du Lézard ocellé en testant la corrélation entre différentes variables environnementales et la probabilité de présence de l'espèce.



## Méthodes

### Principe des modèles de distribution :

- Modéliser la probabilité de présence relative d'une espèce en fonction de covariables environnementales spatialisées.
- Projeter de manière spatialement explicite les prédictions du modèle précédent.

### Données nécessaires :

- Observations de Lézards ocellés géo-référencées (cf. Figure 3) : nous avons utilisé 3 300 données d'observations géo-référencées (précision inférieure à 100m) collectées sur la période 2007-2016. Ces données proviennent du SINP PACA (SILENE PACA) et de la base de données MALPOLON gérée par l'EPHE.
- Les données utilisées pour ce type de modèle étant des données de présences « seules » (nous n'avons pas de relevés d'absences vraies), il est nécessaire de sélectionner dans la zone d'étude des « pseudo-absences »
- Des covariables spatialisées : nous avons utilisé 9 variables d'habitats, 2 variables topographiques et 3 variables climatiques (voir le détail ci-dessous).

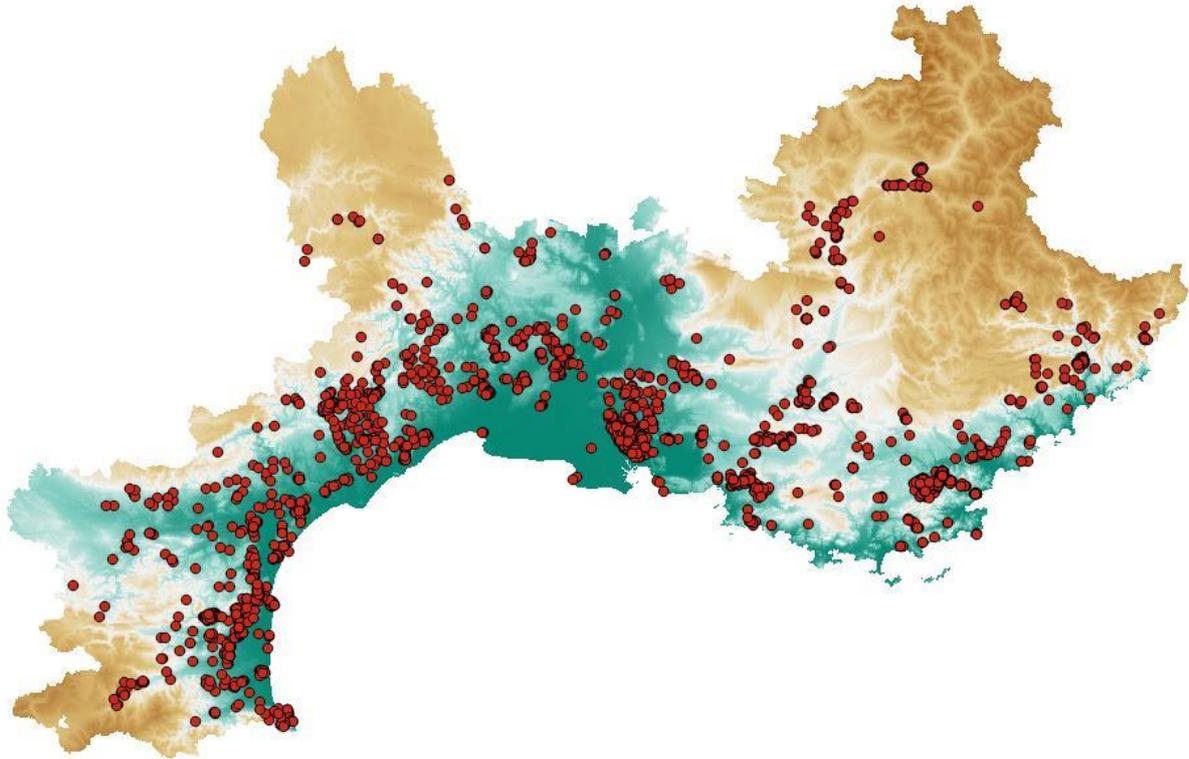


Figure 3 : Localisation des observations de Lézard ocellé de 2007 à 2016 en régions PACA et ex-LR centralisées dans le cadre du PIRA LO et d'après la base de données MALPOLON gérée par l'EPHE. Fond de carte : BDALTI©IGN.



### Traitement des données :

- La modélisation est effectuée à l'échelle d'un hectare (pixels de 100/100m).
- Seule une observation par pixel d'un ha est retenue pour la modélisation afin de réduire l'effet des biais d'échantillonnage (sur-prospection de certaines zones). 2 513 observations sont donc retenues pour les analyses.

### Les variables retenues pour l'analyse :

➤ 9 variables d'**habitat** ont été utilisées en regroupant différents types d'habitats de la base de données d'occupation du sol (ocsol) de 2006 de la région PACA (<http://www.crige-paca.org>) et Languedoc-Roussillon (<http://www.siglr.org/lassociation/la-structure.html>) (cf. Annexe 1 pour voir les regroupements d'habitat effectués). Liste des habitats :

- Urbain dense
- Urbain diffus
- Milieu ouvert
- Milieu semi-ouvert
- Forêts
- Cultures permanentes
- Terres arables
- Milieu agricole hétérogène
- Zones humides

➤ 2 variables **topographiques** :

- Topographie : Indice d'effet de fond de vallée/plateau ou "topographic position index". Les données sources utilisées sont des données d'altitude issues du SRTM à 90m (<https://www2.jpl.nasa.gov/srtm/>).
- Indice d'humidité : concavité/convexité de la topographie ou "topographic wetness index". Les données sources utilisées sont des données d'altitude issues du SRTM à 90m (<https://www2.jpl.nasa.gov/srtm/>).

➤ 3 variables **climatiques** :

- Amplitude thermique : température maximale du mois le plus chaud - température minimale du mois le plus froid (source : WorldClim - <http://www.worldclim.org/>).
- Irradiation potentielle annuelle : calcul de la quantité de chaleur à partir des données d'altitude en intégrant la pente et l'orientation mais la nébulosité n'est pas prise en compte (source : Météo France Modèle Aurelhy interpolé à 100m).
- Degrés jours annuels cumulés (cf Figure 9) (source : Météo France Modèle Aurelhy interpolé à 100m).

Un certain nombre de variables, notamment topographiques et climatiques n'ont pas été testée du fait de leur corrélation trop importante et ainsi générant un risque de surparamétrisation du modèle (cf. Annexe 2 pour visualiser les résultats des tests de corrélations deux à deux).



### Modèles utilisés :

- Modèles Additifs Généralisés (GAM).
- Un total de 100 GAMs a été ajusté. En effet, 10 tirages aléatoires différents de 2 513 pseudo-absences ont été réalisés (même nombre de pseudo-absence que de présence retenues). Pour chacun de ces tirages, 10 tirages aléatoires de 70% et 30% des données d'observations ont été réalisés.
- Pour chacun des 100 modèles, les 70% des données sont utilisées pour estimer la probabilité de présence relative de l'espèce (calibration du modèle) et 30% pour évaluer la qualité des estimations (validation du modèle).
- La carte de probabilité de présence relative finale a été obtenue par un consensus de l'ensemble des 100 cartes. Ce consensus a été obtenu en calculant la moyenne des prédictions de ces 100 modèles, pondérées par le score obtenu à l'étape de validation de chaque modèle. Le fait de réaliser 100 modèles sur des jeux de données différents permet de quantifier la stabilité des résultats (variabilités des prédictions) et donc leur précision. Cela permet aussi d'identifier les zones pour lesquelles le degré de précision des prédictions est faible (fortes incertitudes) mais également les zones où le degré de précision est fort (faibles incertitudes) à travers le coefficient de variation des prédictions.

### Prise en compte de l'effort d'échantillonnage

L'hétérogénéité de l'effort de prospections est certainement la plus importante source de biais dans la modélisation de ce type de jeu de données car des absences de prospection sont interprétées comme des absences de l'espèce. Afin de minimiser l'impact de l'hétérogénéité spatiale de la prospection sur les prédictions du modèle la démarche a été la suivante :

- pour chaque maille de 100\*100m, il a été calculé le nombre d'espèces de reptiles détectées (hors Lézard ocellé) et le nombre de date d'observations de reptiles (hors Lézard ocellé) des bases de données SILENE (SINP PACA) et de l'EPHE (MALPOLON) et cela dans une zone tampon de 250m autour du centroïde de la maille (afin de considérer une espèce présente dans un périmètre restreint et non seulement à partir d'une donnée ponctuelle). Un total de 43 076 observations précises de reptiles ont ainsi été prises en compte pour ces calculs. tampon de 250m autour de chacune des 43 076 observations a été appliqué.
- Le tirage des pseudo-absences n'a ensuite pas été strictement aléatoire mais, les pseudo-absences sont tirées en priorité dans les mailles avec beaucoup d'espèces détectées et beaucoup de dates d'observations. Cela permet de sélectionner des pseudo-absences dans des zones pour lesquelles on sait qu'il y a eu des efforts de prospections. Les zones moins échantillonnées (sous-prospectées) seront alors prédites, à juste titre, avec un plus faible degré de précision des prédictions.

Plus de détails sur les modèles, leur construction et leur principe, ainsi que des références bibliographiques justifiant les différents choix méthodologiques peuvent être demandé aux auteurs.



## Résultats et discussions

Dans sa globalité, le modèle de distribution potentielle obtenu à l'issue de la modélisation correspond à ce qui pouvait être attendu avec des prédictions de probabilité de présence relative forte au cœur du biome méditerranéen et plus faibles dès lors que l'on approche des biomes montagnards (Alpes, Massif central et Pyrénées) et continentaux.

Au cœur du biome méditerranéen, certaines zones apparaissent avec des probabilités de présence relative assez hétérogènes (comme dans le département du Var par exemple). Cela est probablement dû à la présence de milieux fermés qui ne permettent pas à l'espèce de se maintenir.

Inversement, il est intéressant d'observer que la prédiction de la probabilité de présence relative de l'espèce pénètre, en limite nord de distribution, dans certaines vallées au sein desquelles le Léopard ocellé n'est pas connu et offre alors des perspectives de prospections ciblées intéressantes.

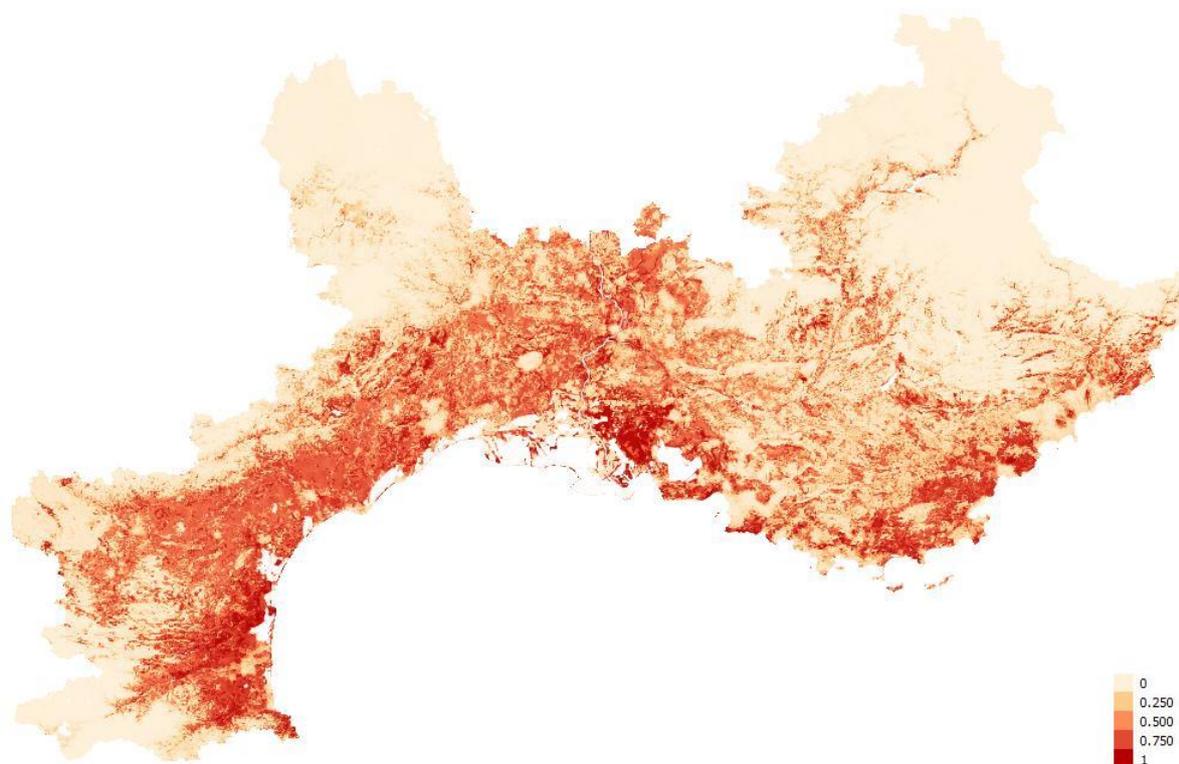
Le modèle de distribution potentielle avec prise en compte de l'effort d'échantillonnages (cf. Figure 4) présente un résultat quasi-similaire au modèle sans prise en compte de l'effort d'échantillonnages (cf. Figure 5). Toutefois, la prise en compte de l'effort d'échantillonnages permet d'améliorer la qualité des prédictions, notamment aux marges de la distribution. La Figure 7 permettent de visualiser les cartes des coefficients de variation, il est en effet visible que les zones où le modèle de distribution est le moins précis sont les zones en limite de distribution connue du Léopard ocellé, où peu d'observations sont disponibles.

**Télécharger les couches raster de la probabilité de présence relative du Léopard ocellé et des coefficients de variations associés via ce lien :**

[www.cen-paca.org/docs/modelisation\\_levard\\_ocelle.zip](http://www.cen-paca.org/docs/modelisation_levard_ocelle.zip)

**Pour accéder à la cartographie de la probabilité de présence relative du Léopard ocellé en ligne, suivre ce lien :** <http://websie.cefe.cnrs.fr/lizmap/>

*\* En cas de disfonctionnement des liens, contacter [marcantoine.marchand@cen-paca.org](mailto:marcantoine.marchand@cen-paca.org)*



*Figure 4 : Probabilité de présence relative du Lézard ocellé avec prise en compte de l'effort d'échantillonnage. De 0 à 1 : probabilité de présence relative faible à forte.*

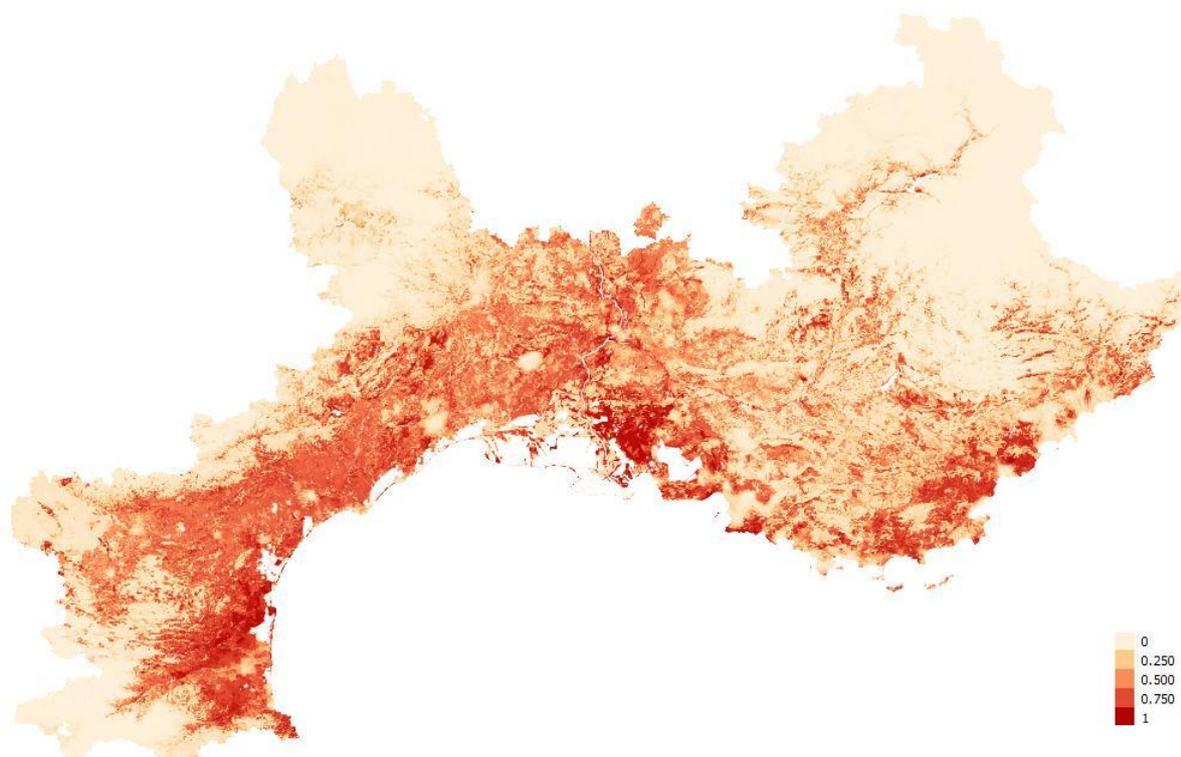


Figure 5 : Probabilité de présence relative du Lézard ocellé sans prise en compte de l'effort d'échantillonnage. De 0 à 1 : probabilité de présence relative faible à forte.

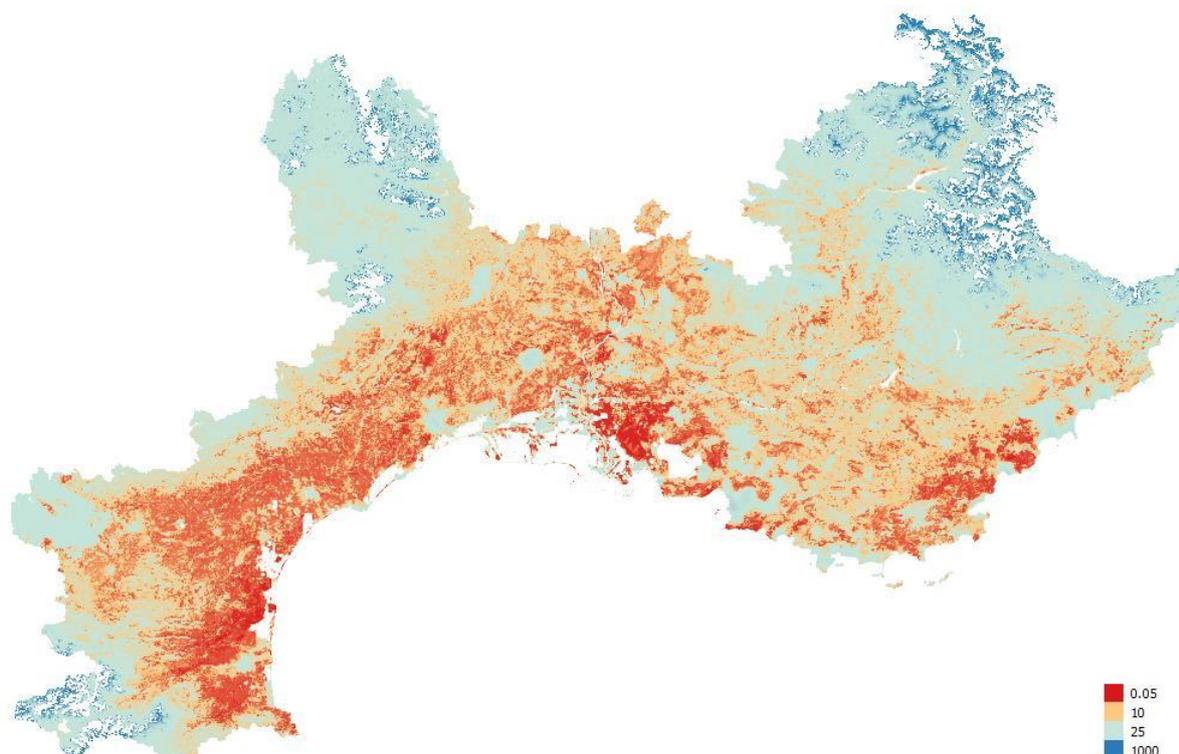


Figure 6 : Coefficient de variation de la probabilité de présence relative du Lézard ocellé sans prise en compte de l'effort d'échantillonnage. Plus la valeur du coefficient de variation est élevée, plus l'incertitude est forte (pour plus de précisions sur la méthode de calcul des coefficients de variations : <https://cran.r-project.org/web/packages/biomod2/biomod2.pdf>).

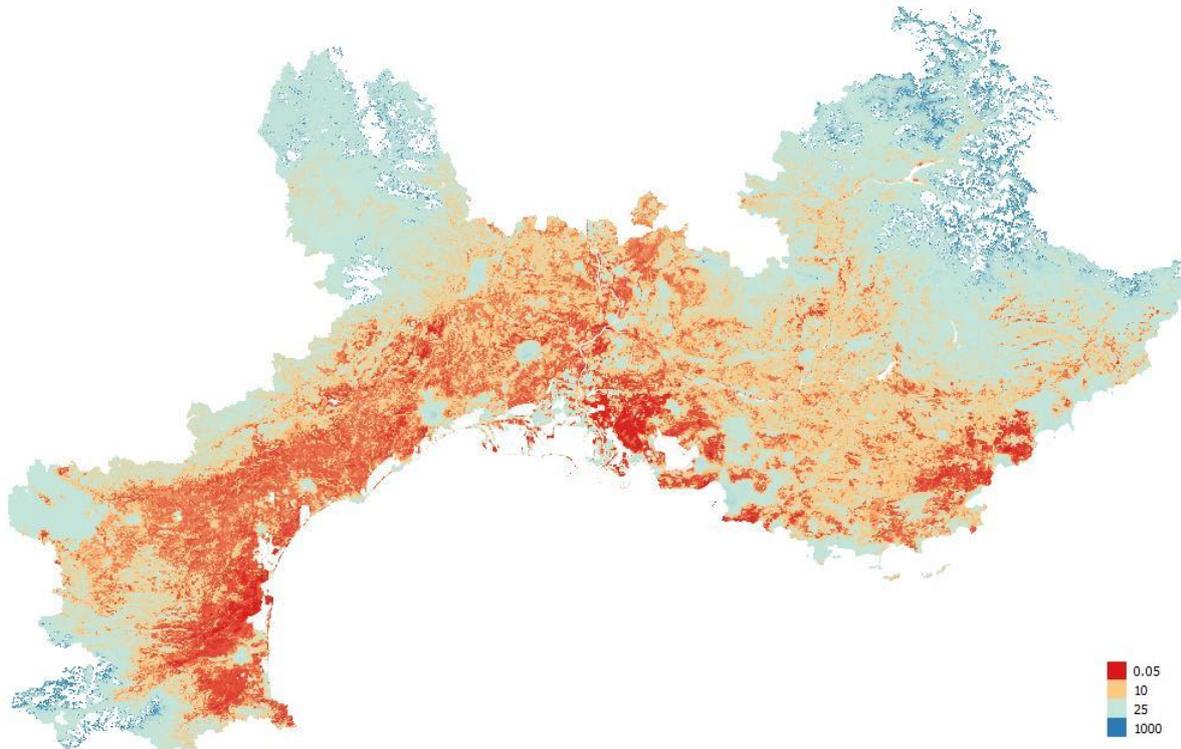


Figure 7 : Coefficient de variation de la probabilité de présence relative du Lézard ocellé avec prise en compte de l'effort d'échantillonnage. Plus la valeur du coefficient de variation est élevée, plus l'incertitude est forte (pour plus de précisions sur la méthode de calcul des coefficients de variations : <https://cran.r-project.org/web/packages/biomod2/biomod2.pdf>).

## Contribution des variables au modèle de distribution potentielle

La variable représentant la température, appelée 'Degrés / jour cumulés' (cf. Figure 9) est la variable qui explique le mieux la présence du Léopard ocellé et est donc celle qui contribue le plus (55%) dans le modèle de distribution (cf. Figure 8). Les habitats ouverts et les habitats forestiers contribuent également bien au modèle avec des valeurs moyennes de 20% et 10%. Les milieux semi-ouverts et l'urbanisation diffuse contribue dans une moindre mesure mais reste importante pour la qualité des prédictions notamment à fine échelle.

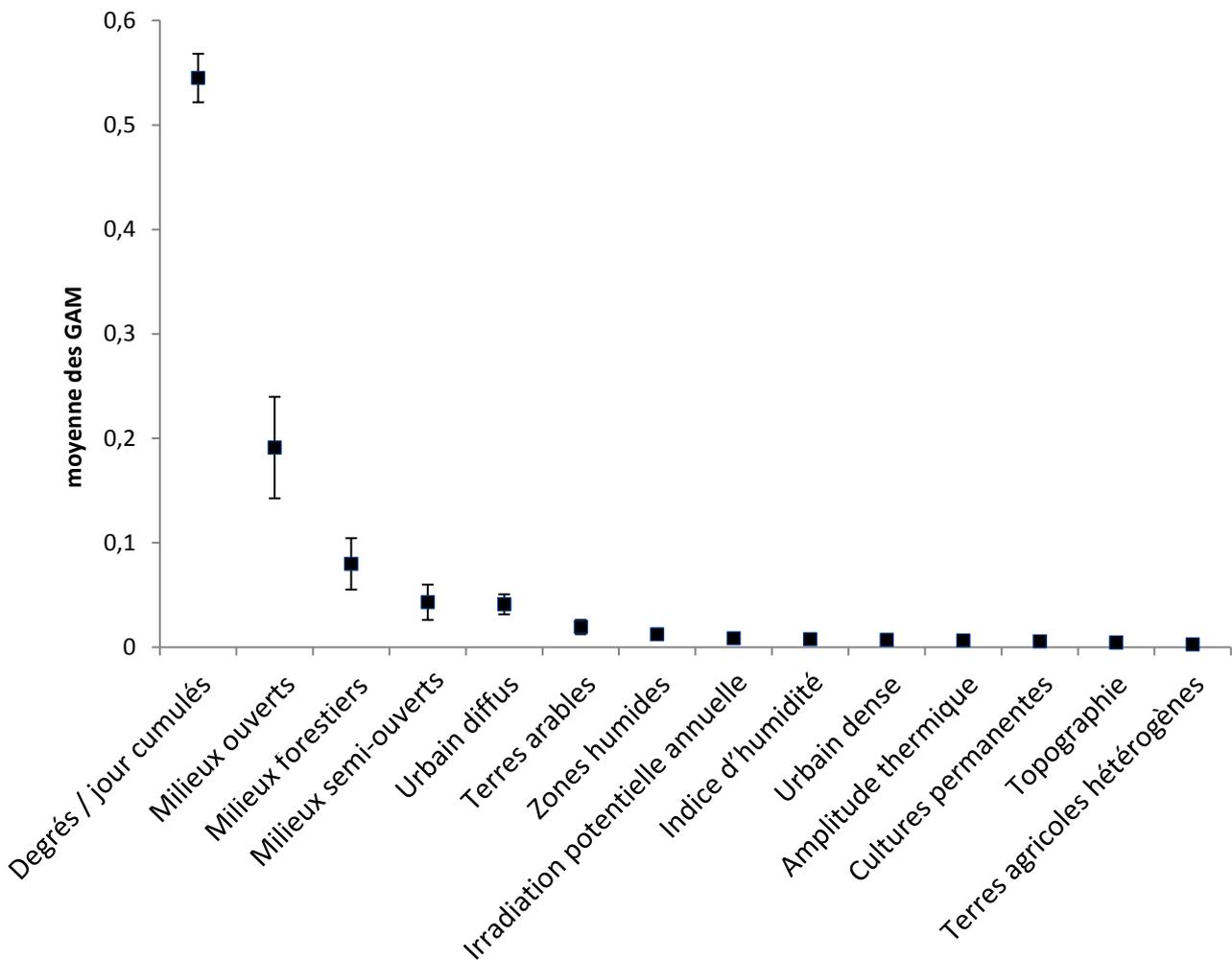


Figure 8 : Contribution moyenne de chaque variable testée au modèle de distribution final. Les boites d'erreurs représentent l'intervalle de confiance à 95% des contributions obtenues sur les 100 modèles.

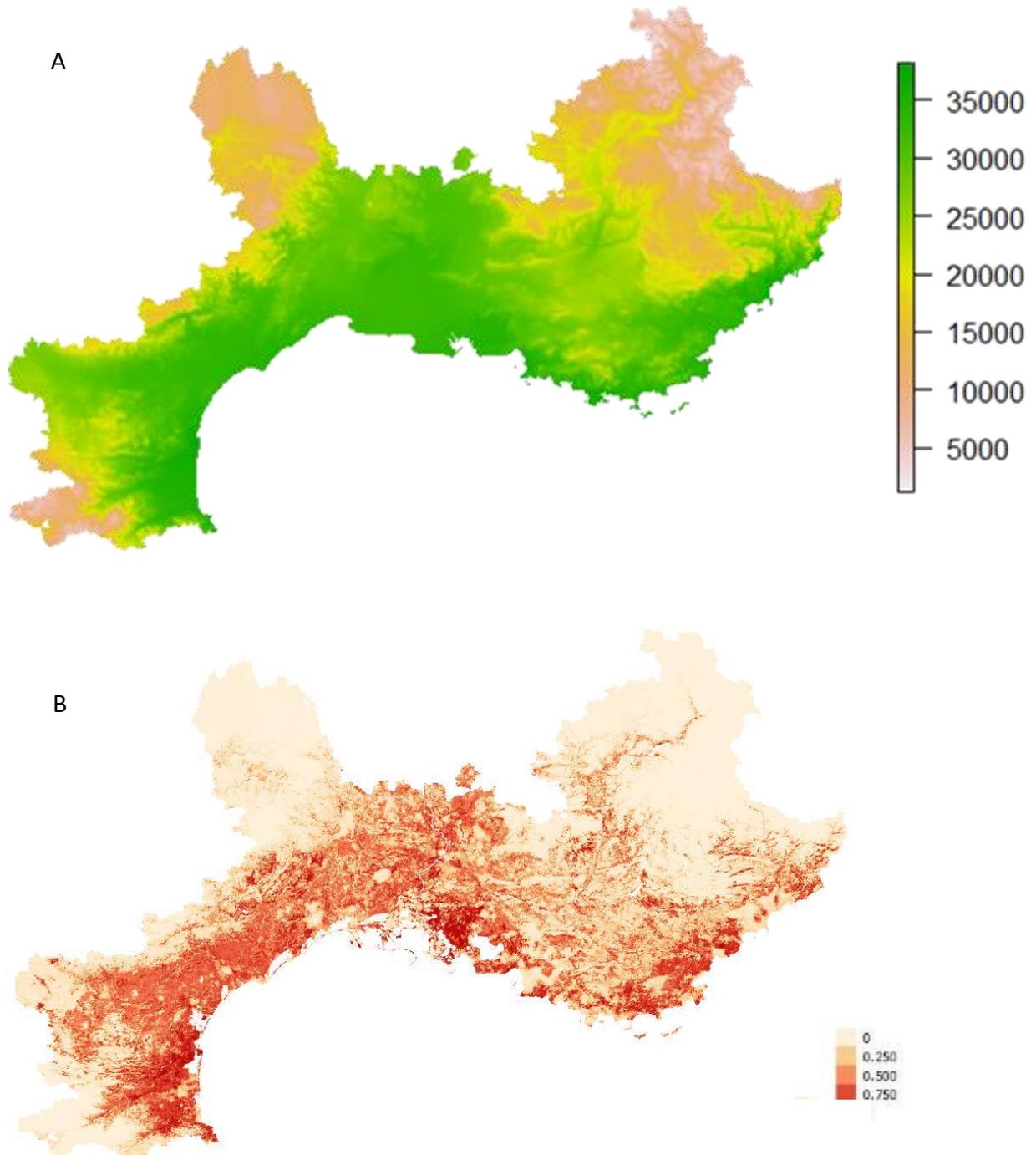


Figure 9 : A : Variable représentant les degrés / jours cumulés. De 0 à 40 000 : température basse à haute. B : Probabilité de présence relative du Lézard ocellé avec prise en compte de l'effort d'échantillonnage. De 0 à 1 : probabilité de présence relative faible à forte.



## Courbes de réponses des variables utilisées pour modéliser la distribution du Lézard ocellé

La visualisation des relations entre la probabilité de présence du Lézard ocellé et les variables qui ont le plus de poids dans le modèle de distribution permet de s'assurer que celles-ci soient cohérentes et puissent être interprétées. Ces graphiques permettent également de bien décrire les relations statistiques entre chaque variable et les probabilités de présence de l'espèce et ainsi de mieux cerner les besoins/contraintes écologique de l'espèce. La Figure 10 permet de visualiser ces relations (qui se nomment « courbes de réponses » dans le jargon des modèles de distribution potentielle).

Les variables qui ont le plus de poids présentent des courbes de réponses lissées, non « erratiques » et très cohérentes avec ce qui pouvait être attendu au regard des connaissances sur l'écologie du Lézard ocellé. En effet, la probabilité de présence de l'espèce est prédite forte sur les zones de hautes températures, dans les milieux ouverts à semi ouverts et faible dans les zones forestières et les zones urbaines. Bien que les autres variables contribuent faiblement au modèle de distribution, certaines sont difficilement interprétables telles que les milieux urbains denses, l'irradiation potentielle annuelle et la topographie. Ceci est lié au fait que leur effet est très variable dans les 100 modèles ajustés, ce qui est lié soit au fait que ces variables ont peu d'effet (donc les courbes de réponses sont très sensibles aux aléa de l'échantillonnage des pseudo-absences et des données dans le jeu de données utilisé pour ajuster les modèles) soit que certaines portions de leurs gradient sont spatialement peu représentées dans la zone d'étude et donc dont l'influence est difficile à prédire (c'est le cas par exemple de l'urbanisation très dense).

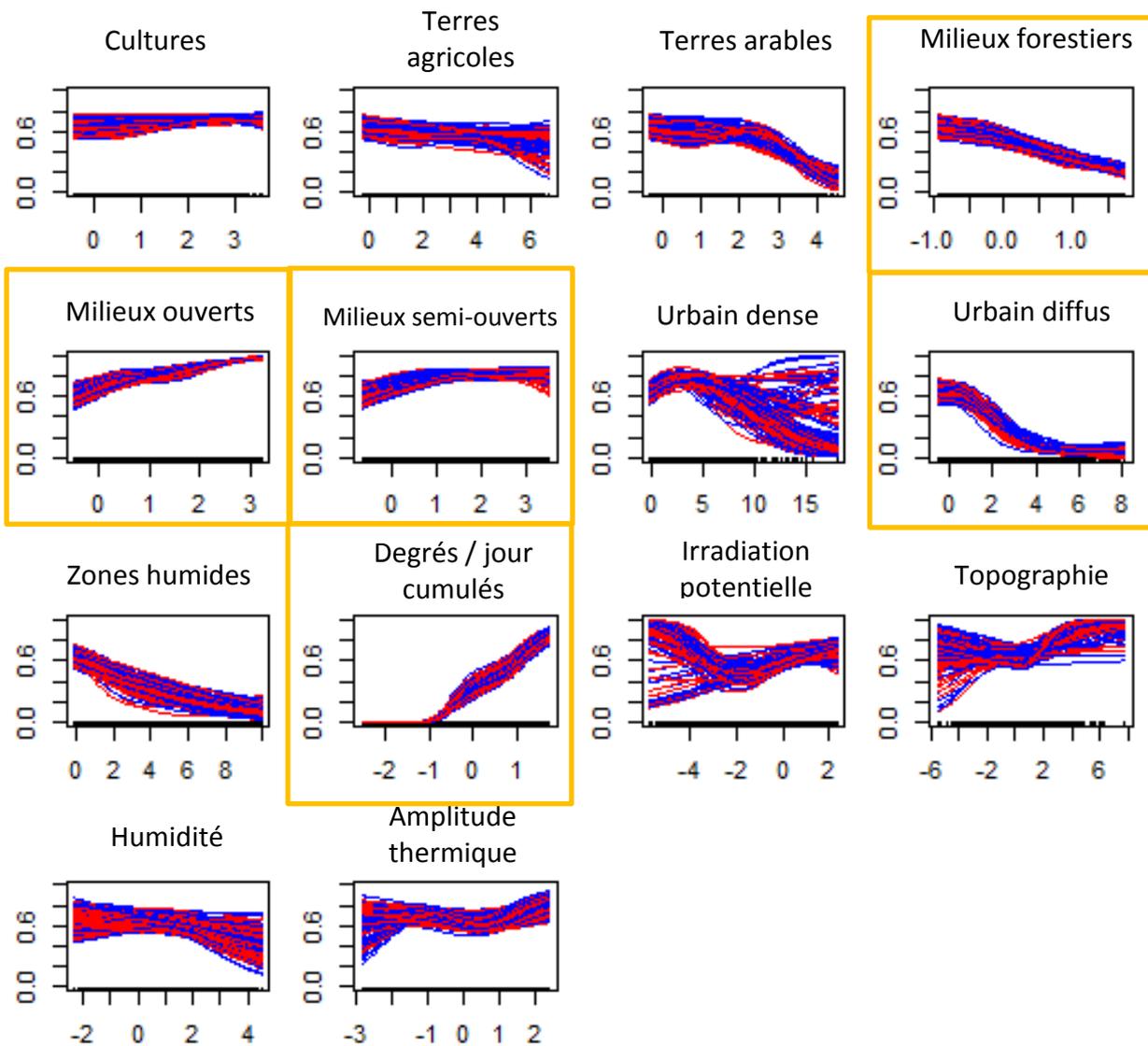


Figure 10 : Courbes de réponses des 14 variables retenues pour prédire la probabilité de présence relative du Lézard ocellé. Pour chaque variable sont représentées les courbes des 100 GAMs générés par les tirages aléatoires différents. Sur chaque graphique, verticalement se lit la probabilité de présence relative du Lézard ocellé (de 0 à 1) et horizontalement la valeur de la variable (de gauche à droite : valeurs faibles à fortes). Encadrés orange : variables ayant le plus de poids dans la distribution du modèle.



## Tests de validation du modèle de distribution

La sensibilité (qui est la capacité à prédire correctement la présence), la spécificité (qui est la capacité à prédire correctement les absences) et la précision globale (qui est la capacité à prédire correctement la présence et l'absence) sont parmi un ensemble des mesures basiques permettant de décrire les liens qui relient les observations (présence/absences) et les prédictions (présence/absences) d'un modèle. Par la combinaison de ces mesures, des méthodes d'évaluations ont été proposées pour juger au mieux de la précision des prédictions des modèles. La Figure 11 ci-dessous illustre les valeurs issues de trois méthodes, parmi les plus performantes et couramment utilisées : le Kappa (ou Kappa de Cohen), le TSS (True Skill Statistic) et le ROC (Relative Operating Characteristic).

Le Kappa permet de pondérer la précision globale des prédictions du modèle par le simple effet du hasard. Ce calcul permet aussi de prendre en compte les prédictions des fausses absences et des fausses présences (erreurs d'omission et de commission). La valeur du Kappa est comprise entre -1 et 1 où 1 correspond à des prédictions parfaites et des valeurs égales ou inférieures à 0 correspondent à des prédictions identiques au hasard. Toutefois, il a été montré que cette méthode est plus ou moins sensible à la prévalence (proportion de cas où l'espèce est considérée comme présente) en fonction de la spécificité et de la sensibilité (Allouche *et al.*, 2006) ce qui rend cette métrique peu enclin à faire des comparaisons robustes de modèles entre espèces, zones d'étude ou jeux de données différents (McPherson *et al.*, 2004). Elle est donc à interpréter avec prudence.

Le TSS proposé par Allouche *et al.* (2006) se veut être une mesure alternative car indépendante de la prévalence du jeu de données et combinant par addition la sensibilité (prédire correctement la présence) et l'inverse de la spécificité (prédire faussement la présence). Ainsi, l'évaluation par la TSS varie entre 0 (prédictions non différentes du hasard) et 1 (prédictions parfaites). Toutefois une étude récente montre des limites dans l'utilisation de cette mesure (Somodi *et al.*, 2017) et notamment les conditions d'utilisation et les hypothèses associées.

Le ROC (Hanley & McNeil, 1982) est une autre méthode qui permet de construire une courbe faisant le lien entre la sensibilité et l'inverse de la spécificité. Cette courbe permet ainsi de calculer la valeur d'AUC (Area Under the ROC Curve). Comme le TSS, le critère ROC permet d'évaluer la capacité du modèle à prédire correctement les présences et les absences de l'espèce indépendamment de la prévalence. L'évaluation par l'AUC varie entre 0.5 (prédictions du modèle non différentes du hasard) et 1 (prédictions parfaites des données observées par le modèle). Tout comme le TSS, l'utilisation de l'AUC possède des limites et notamment dans le cadre de modélisation se basant sur la génération de pseudo-absences (Lobo *et al.*, 2008 ; Jiménez-Valverde, 2012)

Les prédictions sont jugées pertinentes dès lors qu'elles correspondent à une valeur d'AUC > 0.7 ou une valeur de TSS > 0.4 en sachant que dans la pratique les deux évaluations sont fortement corrélées (Swets, 1988 ; Engler *et al.*, 2011).

Dans le cas de notre travail, le Kappa est de 0.60, le TSS est de 0.60 et l'AUC est de 0,88, ce qui peut être considéré comme une très bonne évaluation du modèle à prédire la répartition de l'espèce. Si une évaluation égale à 1 peut paraître idéal en théorie, en pratique il est généralement signe d'une sur-paramétrisation du modèle et d'une faible capacité à prédire la



répartition de l'espèce en dehors des zones de présences et d'absences connues de l'espèce. En effet, aucun modèle, aussi réaliste qu'il le soit ne peut prédire avec certitude les présences et les absences d'une espèce dans l'espace.

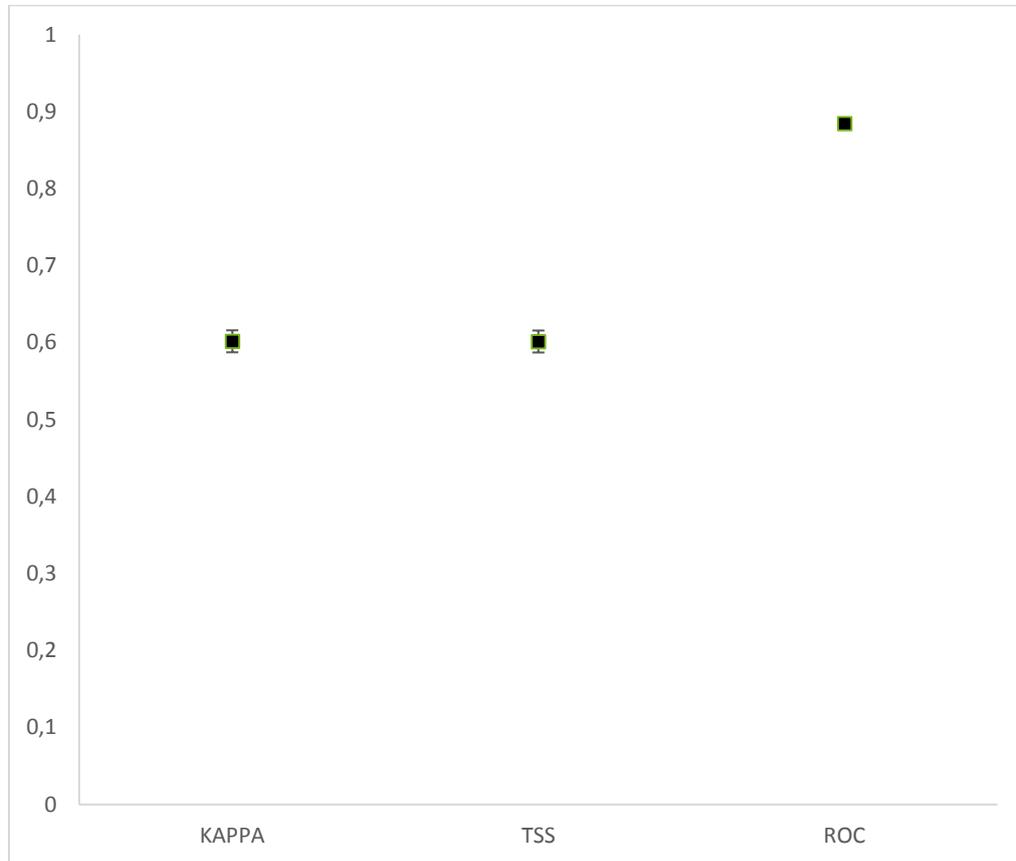


Figure 11 : Moyenne et intervalle de confiance à 95% des valeurs des tests de validation des modèles de distribution sur les 100 modèles ajustés.

### Confrontation des données d'observation au modèle de distribution

La superposition spatiale des prédictions et des données d'observations de Lézard ocellé sur certaines zones permet également de vérifier dans le détail la qualité de prédiction du modèle.

Dans les secteurs situés en limite de distribution, les prédictions semblent moins pertinentes par rapport à ce qui est connu en zone cœur. C'est le cas au sud du département de la Lozère et au nord du département du Gard (cf. Figure 12).

Cela peut venir du fait que l'on ne peut pas intégrer toutes les composantes qui expliquent la répartition d'une espèce (historique, subtilités locales non intégrées dans les variables, micro conditions, observations douteuses / individus isolés,...). Cela peut également venir du fait que pour les individus en marge de leur niche d'autres facteurs peuvent prendre un poids plus important pour expliquer leur persistance/absence.

Comme énoncé plus haut, ces secteurs en limite de distribution sont les endroits où la modélisation à le plus de difficultés à prédire la probabilité de présence relative du Lézard ocellé (fort coefficient de variation, cf. Figure 4), ce sont de fait les endroits pour lesquels des efforts de prospections ciblées pourraient être intéressants.

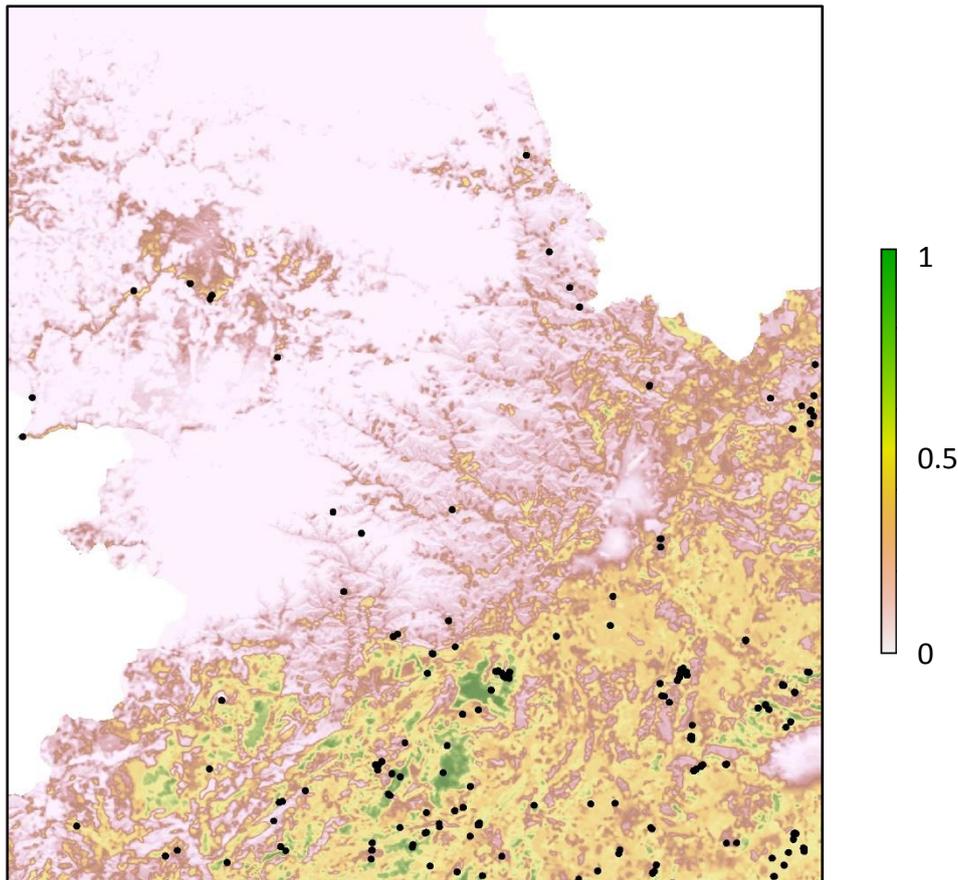


Figure 12 : Superposition du modèle de distribution potentielle et des données d'observations au sud de la Lozère et au nord du Gard. Points noirs : observation de Lézard ocellé ; de 0 à 1 : probabilité de présence relative faible à forte.

Au cœur de l'aire de distribution connue, les données d'observations dont nous disposons se superposent correctement sur les zones dont la probabilité de présence est forte. C'est le cas sur le secteur s'étalant de la plaine de Maures jusqu'au Préalpes d'Azur en passant par le massif de l'Esterel (cf. Figure 13). A noter que des secteurs qui apparaissent comme très favorables ne présentent aucune observation. Ces secteurs peuvent être intéressants à prospecter pour valider la qualité du modèle et l'améliorer.

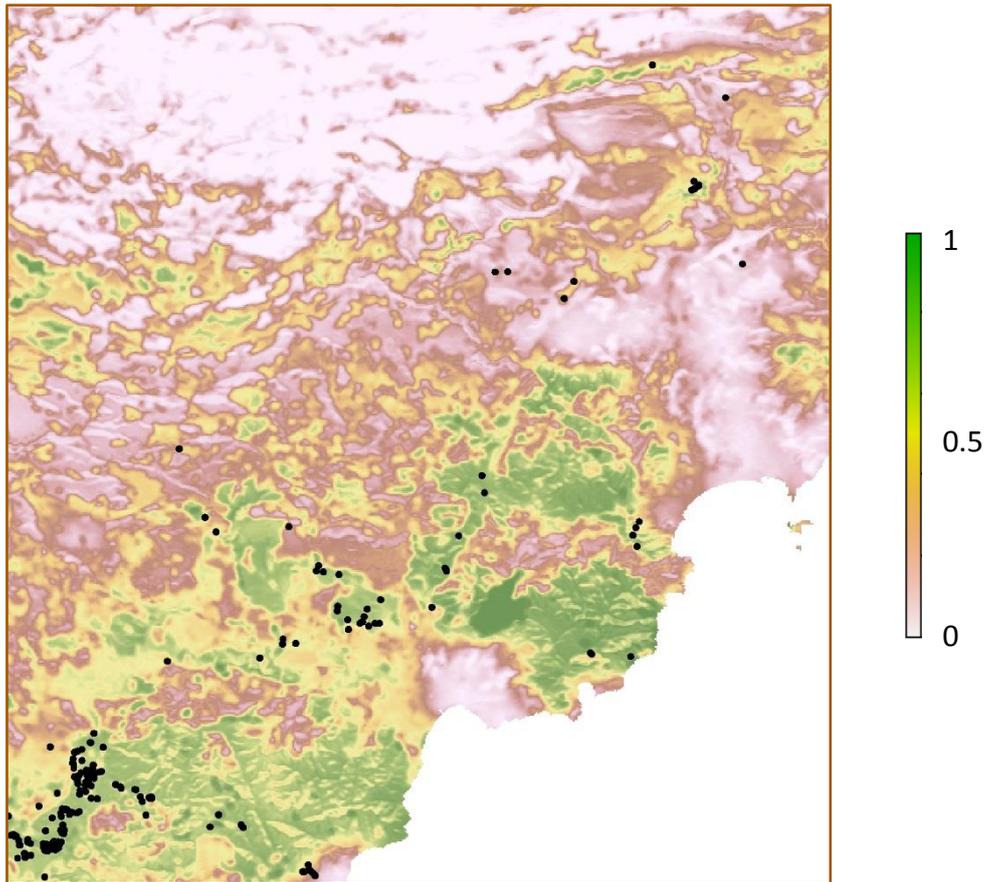


Figure 13 : Superposition du modèle de distribution potentielle et des données d'observations sur le secteur de la plaine des Maures jusqu'au Préalpes d'Azur. Points noirs : observation de Lézard ocellé ; de 0 à 1 : probabilité de présence relative faible à forte.



## Interprétation et précautions d'utilisation de l'outil

Pour ce type de méthode d'analyses, l'hétérogénéité spatiale des efforts de prospections (accessibilité des sites, pente,...) est probablement le facteur le plus limitant à la réalisation d'un modèle de distribution potentiel fiable. C'est pourquoi un indice d'effort de prospection a été intégré aux analyses (nombre d'espèces et nombre de dates d'observations) afin d'atténuer ce biais. Cependant, il n'est pas sûr que cette solution soit suffisante pour corriger l'ensemble des biais.

Les bases de données ocsol utilisées pour créer les variables d'habitats datent de 2006. Les données d'observations utilisées pour le modèle datent de 2007 à 2016. Ainsi, les incohérences temporelles sont minimisées. En effet, une observation effectuée avant la définition d'un habitat qui aurait évolué entre la date de l'observation et la date de définition de l'habitat (e.g. : garrigue transformée en zone commerciale) aurait pu biaiser les analyses. Toutefois, des imprécisions spatiales, mais aussi de typologie d'habitat, demeurent dans ces bases de données réalisées à partir d'images satellites. Par ailleurs, certaines zones ont pu évoluer localement.

Le Lézard ocellé est une espèce dont la dépendance aux gîtes a été démontrée par plusieurs études (Grillet *et al.*, 2010 ; Bourgault, 2013 ; Salgado & Hernandez, 2013). Or la présence de gîtes est une information impossible à obtenir sur une large échelle spatiale et donc impossible à prendre en compte dans le cadre de ce travail. Ceci est une limite forte à ce type d'approche car l'espèce peut être prédite comme probablement présente dans des zones avec une faible capacité d'accueil du fait de manque de gîtes. Toutefois, des mesures de restaurations/créations de gîtes peuvent être priorisées dans ces mêmes secteurs.

Comme souvent avec ce type de modèles les prédictions sont moins précises aux marges de la distribution de l'espèce (limites nord). Ceci est lié au fait que les pixels dans ces parties de gradients sont moins nombreux et que les données d'observation y sont aussi généralement plus rares (soit parce que l'espèce y est moins bien recherchée, soit parce qu'elle occupe ces zones en plus faible abondance donc y est plus rarement détectée), soit parce que la limite de répartition de l'espèce n'est pas une limite franche mais plutôt un émiettement progressif des zones favorables à l'espèce.

Cet outil doit donc être considéré comme une aide à la prise en compte du Lézard ocellé dans certaines études et non comme une cartographie à prendre telle quelle. La présence ou l'absence du Lézard ocellé ne peut être certifiée par ce modèle de distribution seul car il s'accompagne de sources de biais et d'incertitude comme tout modèle (« tous les modèles sont faux, certains sont utiles », George Edward Pelham Box). Il doit donc être accompagnée 1/ d'appréciations complémentaires de l'habitat telles que la présence ou l'absence de gîtes et 2/ de prospections ciblées sur l'espèce. Il convient donc d'utiliser cet outil avec précaution et de ne pas négliger les prospections de terrain.



## Perspectives

Une approche mécaniste sera développée en parallèle de cette modélisation. Une des pistes envisagées est de se baser sur les contraintes de l'espèce à se reproduire du fait de l'importance des conditions thermiques nécessaires au bon développement des embryons, dans les sites de pontes. Cela nécessitera de définir expérimentalement les liens entre la température et la durée d'incubation des œufs mais aussi de prédire les conditions potentielles d'incubation à l'échelle de l'aire de répartition de l'espèce en France. L'hypothèse est que la date d'éclosion doit avoir un effet fort sur le succès de reproduction et par conséquent sur le maintien des populations. Prédire les dates d'éclosion potentielles de l'espèce à l'échelle de sa répartition permettra d'explorer spatialement ce facteur limitant et d'identifier par exemple les régions les plus sensibles *a priori* à la fermeture des milieux. De plus, la création d'une nouvelle variable spatiale qui représenterait la durée d'incubation potentielle pourrait également être intégrée aux modélisations de distribution potentielle pour prédire plus finement la distribution actuelle de l'espèce. Les modèles mécanistes permettent également d'exprimer spatialement des scénarios comme les effets de la fermeture et de réouverture des milieux sur la durée d'incubation et *in fine*, leurs conséquences sur les marges de répartition de l'espèce par l'intermédiaire des modèles de niches.

La mise en œuvre des inventaires et suivis par placettes selon le protocole standardisé mis en place dans le cadre du PIRA apporte une approche complémentaire avec la prise en compte notamment de la faible détection du Lézard ocellé pour prédire la probabilité de présence, absolue et non relative cette fois, de cette espèce. Toutefois, le plan d'échantillonnage spatial se doit d'être spatialement homogène afin d'avoir une approche représentative de la zone d'étude (ce qui n'est pas le cas pour le moment). Dans les perspectives il est en théorie possible de combiner les données du protocole standardisé et celles de bases de données occasionnelles. Ce type de combinaison est relativement complexe et sera entreprise dans les années à venir.

En attendant ces futurs développements, nous ferons évoluer ce modèle au fur et à mesure que de nouvelles données seront disponibles, nous engageons donc les observateurs à renvoyer les données (prospections positives et aussi négatives) mais aussi à cibler des zones de prospection à partir du modèle (zones en marges pour les préciser, zones prédites comme de forte probabilité de présence sans observation, etc.). Ces efforts ciblés sont susceptibles de conduire à une amélioration rapide de la qualité des prédictions et ainsi encore mieux cerner la répartition effective (découverte de nouvelle population/ vérification de l'absence) mais aussi la répartition potentielle (restauration des milieux/corridor pour que l'espèce puisse occuper pleinement les secteurs favorables). **Pour accéder à la cartographie de la probabilité de présence relative du Lézard ocellé dans les régions Provence-Alpes-Côte d'Azur et ex-Languedoc-Roussillon, suivre ce lien : <http://websie.cefe.cnrs.fr/lizmap/>**



## Bibliographie

- Allouche O, Tsoar A & Kadmon R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* 43 : 1223-1232.
- Beven K & Kirkby MJ. 1979. A physically-based variable contribution area model of catchment hydrology. *Hydrology Science Bulletin* 24 : 43-69.
- Bourgault L. 2013. Etude de la dynamique spatio-temporelle du Léopard ocellé (*Timon lepidus*, Daudin, 1802) par radio-téléométrie sur la Chaîne de l'Etoile et le Massif du Garlaban. *Colinéo*. 55p.
- Doré F & Thirion JM. 2012. Plan national d'actions Léopard ocellé *Timon lepidus* 2012 – 2016. *Ministère de l'écologie, Obios*. 114 p.
- Engler R, Randin CF, Thuiller W, Dullinger S, Zimmermann NE, Araujo MB, Pearman PB, Le Lay G, Piedallu C, Albert CH, Choler P, Coldea G, De Lamo X, Dirnbock T, Gegout JC, Gomez-Garcia D, Grytnes JA, Heegaard E, Hoistad F, Nogues-Bravo D, Normand S, Puscas M, Sebastia MT, Stanisci A, Theurillat JP, Trivedi MR, Vittoz P & Guisan A. 2011. 21st century climate change threatens mountain flora unequally across Europe. *Global Change Biology* 17 : 2330-2341.
- Grillet P, Cheylan M, Thirion JM, Doré F, Bonnet X, Dauge C, Chollet S & Marchand MA. 2010. Rabbit burrows or artificial refuges are a critical habitat component for the threatened lizard, *Timon Lepidus* (Sauria, Lacertidae). *Biodiversity and Conservation* 19 : 2039-2051.
- Hanley JA & McNeil BJ. 1982. The meaning and use of the area under a Receiver Operating (ROC) Curve characteristic. *Radiology* 143 : 29-36.
- Jiménez-Valverde A. 2012. Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Global Ecology and Biogeography* 21 (4) : 498-507.
- Legouez C & Marchand MA. 2013. Plan interrégional d'actions de Provence-Alpes-Côte d'Azur et du Languedoc-Roussillon en faveur du Léopard ocellé *Timon lepidus* (Daudin, 1802) 2013 - 2017. *Conservatoire d'espaces naturels de PACA*. 130 p.
- Lobo JM, Jiménez-Valverde A & Real R. 2008. AUC : a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* 17 (2) : 145-151.
- McPherson JM, Jetz W & Rogers DJ. 2004. The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact ? *Journal of Applied Ecology* 41 : 811-823.
- Salgado I & Hernandez MA. 2013. *Timon lepidus* usa las conejas como refugio. *Bol. Asoc. Herpetol. Esp.* 24 (2) : 59-61.
- Somodi I, Lepesi N & Botta-Dukát Z. 2017. Prevalence dependence in model goodness measures with special emphasis on true skill statistics. *Ecology and Evolution* 7 (3) : 863-872.
- Swets JA. 1988. Measuring the accuracy of diagnostic systems. *Science* 240 : 1285-1293.



## Annexes

Annexe 1 : regroupements des habitats effectués à partir des bases de données d'occupation du sol des régions PACA et ex-LR.

habitats des bases de données d'occupation du sol des régions PACA et LR			regroupements effectué pour la modélisation
Intitulé Niv1	Intitulé Niv2	Intitulé Niv3	
Territoires artificialisés	Zones urbanisées	Tissu urbain continu	Urbain dense
		Tissu urbain discontinu	Urbain diffus
		Bâti diffus	Urbain diffus
	Zones industrielles et commerciales et réseaux de communication	Zones industrielles ou commerciales	Urbain dense
		Réseaux routier et ferroviaire et espaces associés	Urbain diffus
		Zones portuaires	Urbain dense
		Aéroports	Urbain dense
	Mines, décharges et chantiers	Extraction de matériaux	Urbain dense
		Décharges	Urbain dense
		Chantiers	Urbain dense
Espaces verts artificialisés, non agricoles	Espaces verts urbains	Urbain diffus	
	Equipements sportifs et de loisirs	Urbain diffus	
Territoires agricoles	Terres arables	Serres	Terres arables
		Terres arables autres que serres, zones à forte densité de serres et rizières	Terres arables
		Rizières	Terres arables
		Zones à forte densité de serres	Terres arables
	Cultures permanentes	Vignobles	Cultures permanentes
		Vergers et petits fruits	Cultures permanentes
		Oliveraies	Cultures permanentes
		Lavandins	Cultures permanentes
Prairies	Prairies	Milieu ouvert	

habitats des bases de données d'occupation du sol des régions PACA et LR			regroupements effectués pour la modélisation
Intitulé Niv1	Intitulé Niv2	Intitulé Niv3	
Territoires agricoles	Zones agricoles hétérogènes	Cultures annuelles associées aux cultures permanentes	Terres agricoles hétérogènes
		Systèmes culturaux et parcellaires complexes	Terres agricoles hétérogènes
		Territoires principalement composés par l'agriculture avec présence de végétation	Terres agricoles hétérogènes
		Territoires agro-forestiers	Terres agricoles hétérogènes
Forêts et milieux semi-naturels	Forêts	Forêts de feuillus	forêt
		Forêts de conifères	forêt
		Forêts mélangées	forêt
	Milieux à végétation arbustive et/ou herbacée	Pelouses et pâturages naturels	Milieu ouvert
		Landes subalpines	Milieu semi-ouvert
Maquis et garrigue		Milieu semi-ouvert	
Forêt et végétation arbustive en mutation	Milieu semi-ouvert		
Landes	Milieu semi-ouvert		
Zones humides	Zones humides intérieures	Autres zones humides et intérieures	Zones humides
		Marais intérieurs et roselières	Zones humides
		Tourbières	Zones humides
	Zones humides maritimes	Marais maritimes	Zones humides
		Marais salants	Zones humides

Annexe 2 : résultats des tests de corrélations entre les covariables environnementales à tester. Les valeurs colorées sont les variables considérées corrélées et donc qui ne peuvent être intégrées au sein du même modèle. Les variables retenues pour réaliser les modèles sont marquées avec un \*, celles exclues avec un \$.

	Cultures permanentes *	Terres agricoles hétérogènes *	Terres arables *	Forêts *	Milieu ouvert *	Milieu semi-ouvert *	Urbain *	Altitude \$	Pente \$	Degrés / jour cumulés *	Bilan hydrique \$	Irradiation potentielle annuelle *	Topographie *	Indice d'humidité *	Température maximum du mois le plus chaud \$	Température minimale du mois le plus froid \$	Amplitude thermique *	Température minimum en hiver \$	Température minimum en été \$	
Cultures permanentes *	1																			
Terres agricoles hétérogènes *	-0,03	1																		
Terres arables *	-0,07	-0,04	1																	
Forêts *	-0,28	-0,17	-0,22	1																
Milieu ouvert *	-0,11	-0,05	-0,07	-0,21	1															
Milieu semi-ouvert *	-0,15	-0,1	-0,13	-0,29	-0,09	1														
Urbain *	-0,05	-0,01	-0,04	-0,22	-0,08	-0,14	1													
Altitude \$	-0,31	-0,1	-0,16	0,14	0,24	-0,01	-0,21	1												
Pente \$	-0,29	-0,16	-0,23	0,32	0,06	0,01	-0,19	0,65	1											
Degrés / jour cumulés *	0,35	0,07	0,13	-0,21	-0,25	0,01	0,24	<b>-0,94</b>	-0,58	1										
Bilan hydrique \$	-0,32	-0,09	-0,15	0,25	0,2	-0,05	-0,23	<b>0,93</b>	0,67	<b>-0,94</b>	1									
Irradiation potentielle annuelle *	0,08	0,05	0,06	-0,25	0,06	0,12	0,07	-0,17	-0,26	0,17	-0,4	1								
Topographie *	-0,05	-0,03	-0,04	0,05	-0,03	0,03	-0,07	0,15	0,1	-0,11	0,09	0,06	1							
Indice d'humidité *	0,33	0,18	0,27	-0,42	-0,02	-0,13	0,2	-0,56	<b>-0,71</b>	0,53	-0,6	0,16	-0,36	1						
Température maximum du mois le plus chaud \$	0,31	0,11	0,15	-0,13	-0,24	0,02	0,18	<b>-0,97</b>	-0,64	<b>0,92</b>	<b>-0,9</b>	0,19	-0,02	0,51	1					
Température minimale du mois le plus froid \$	0,3	0,07	0,11	-0,12	-0,24	0,06	0,19	<b>-0,92</b>	-0,57	<b>0,93</b>	<b>-0,9</b>	0,17	-0,01	0,43	<b>0,92</b>	1				
Amplitude thermique *	0,19	0,13	0,16	-0,09	-0,12	-0,05	0,06	-0,58	-0,45	0,43	-0,5	0,13	-0,03	0,4	0,65	0,31	1			
Température minimum en hiver \$	0,3	0,07	0,11	-0,12	-0,24	0,06	0,19	<b>-0,93</b>	-0,57	<b>0,94</b>	<b>-0,9</b>	0,17	-0,01	0,44	<b>0,92</b>	<b>1</b>	0,32	1		
Température minimum en été \$	0,31	0,07	0,13	-0,14	-0,26	0,03	0,2	<b>-0,95</b>	-0,58	<b>0,97</b>	<b>-0,9</b>	0,17	-0,01	0,47	<b>0,95</b>	<b>0,98</b>	0,42	<b>0,98</b>	1	

## Modélisation spatiale de la probabilité de présence relative du Lézard ocellé dans les régions Provence-Alpes-Côte d'Azur et ex-Languedoc-Roussillon

### Déclinaison Inter-Régional PACA et LR du PNA Lézard ocellé



**Ecole Pratique des Hautes Etudes**  
46, rue de Lille - 75007 Paris



**Centre d'Ecologie Fonctionnelle et Evolutive UMR 5175**  
1919 Route de Mende - 34293 Montpellier  
[Aurelien.BESNARD@cefe.cnrs.fr](mailto:Aurelien.BESNARD@cefe.cnrs.fr)  
[Guillelme.ASTRUC@cefe.cnrs.fr](mailto:Guillelme.ASTRUC@cefe.cnrs.fr)



**Centre d'Etude Biologique de Chizé UMR 7372**  
405 route de la Canauderie - 79360 Villiers en Bois  
[olivier.lourdais@cebc.cnrs.fr](mailto:olivier.lourdais@cebc.cnrs.fr)  
[guillon.m.g@gmail.com](mailto:guillon.m.g@gmail.com)



**Conservatoire d'espaces naturels de Provence-Alpes-Côte d'Azur**  
96 rue Droite - 04200 Sisteron  
[marcantoine.marchand@cen-paca.org](mailto:marcantoine.marchand@cen-paca.org)

Avec le soutien de :

