

# Construction de cartes d'habitat potentiel

## Comparaison et évaluation de l'ENFA et du GLM à l'aide d'une espèce virtuelle : le marsupilami

Véronique Helfer  
et  
François Métral

sous la direction d'Alexandre H. Hirzel



Travail de certificat - Module d'Ecologie  
Institut de Zoologie et d'Ecologie Animale  
Université de Lausanne

## Table des matières

<b>Introduction</b>	1
<b>Matériel et méthodes</b>	3
<b>A. Construction du modèle</b>	3
<b>Création de la carte de qualité d'habitat</b>	3
1) <i>Choix des variables</i>	
2) <i>Préparation des cartes pour chaque variable</i>	
3) <i>Transformation des variables en "Indices de qualité d'habitat"</i>	
4) <i>Création de la carte de qualité d'habitat</i>	
<b>Répartition des trois sous-espèces</b>	6
1) <i>M. numericus ssp. helveticus</i>	
2) <i>M. numericus ssp. musca-cerevisiae</i>	
3) <i>M. numericus ssp. aliensis</i>	
<b>B. Analyses de la niche</b>	8
<b>Observations sur le terrain</b>	9
- <i>Points de présence</i>	
- <i>Points d'absence</i>	
<b>L'analyse factorielle de la niche écologique (ENFA)</b>	9
<b>Le modèle linéaire généralisé (GLM)</b>	10
<b>C. Validation des modèles</b>	11
- <i>Comparaisons avec la carte d'habitat de référence</i>	
- <i>Evaluation des modèles par les cartes de validation</i>	
<b>Résultats</b>	12
<b>Cartes obtenues avec l'ENFA et le GLM</b>	12
<b>Variables retenues par les deux analyses</b>	14
- <i>L'analyse factorielle de la niche écologique</i>	
- <i>Le modèle linéaire généralisé</i>	
<b>Validation des modèles</b>	16
<b>Discussion</b>	16
<b>Remerciements</b>	16
<b>Bibliographie</b>	20
<b>Annexes</b>	

## Introduction

Les cartes de distribution des espèces animales sont utiles pour voir l'évolution de leur répartition et savoir comment elles se portent : si elles sont en phase de colonisation ou si elles occupent déjà la totalité du territoire, ... Des cartes de répartition vraiment objectives, suffisamment précises et complètes, sont particulièrement difficiles à obtenir pour les espèces difficiles à observer. C'est pourquoi les cartes de domaine potentiel, représentant l'aire sur laquelle la présence de l'espèce est théoriquement possible, sont souvent utilisées. Ces cartes peuvent en outre, au vu des données d'observation, orienter des recherches et stimuler des investigations dans les zones favorables à l'espèce qui ne semblent pas occupées. Les cartes de domaine potentiel sont utiles non seulement en tant que base de données mais aussi dans une optique de gestion des espèces menacées. En effet, à partir d'une carte de domaine potentiel et d'une carte relevant les sites où l'espèce est présente, on pourra repérer quelles sont les zones favorables à l'espèce qui ne sont pas exploitées et proposer un plan de gestion permettant à l'espèce de coloniser de tels endroits par la construction de corridors à faune, par exemple. On peut aussi déterminer des zones moins favorables et proposer divers aménagements susceptibles de les améliorer. L'utilité de ces cartes est multiple, mais encore faut-il savoir si l'on peut vraiment s'y fier.

Les différentes méthodes d'analyse permettant d'aboutir à de telles cartes se basent sur les données d'observations et relèvent différentes variables écogéographiques (climatiques, topographiques, végétales, d'occupation du territoire) permettant de qualifier les sites d'observation. On peut retenir l'analyse discriminante (Zimmermann, 1998), la régression logistique (Akçakaya, 1997 ; Manel & al., 1999), qui peuvent se ramener à des réponses environnementales linéaires, le réseau neuronal artificiel (Manel & al., 1999) et le modèle linéaire généralisé (Maggini, 1999), qui tiennent compte d'interactions plus complexes. Pour cartographier la qualité de l'habitat, il faut travailler à une échelle relativement fine qui permet d'obtenir des zones suffisamment homogènes par rapport aux variables écogéographiques et d'assurer que ces dernières soient assez précises. Une grille est superposée à la zone étudiée, chaque quadrat ainsi défini aura une valeur dépendante des variables écogéographiques relevées.

Toutes les méthodes citées ci-dessus font appel à des données de présence et d'absence, pour lesquelles elles comparent les valeurs des variables écogéographiques. Si elles sont bien adaptées à la modélisation des habitats favorables à différentes espèces végétales, elles posent problème pour les analyses concernant le règne animal. En effet, les données d'absence ne sont pas fiables pour les organismes animaux car le fait qu'on ne les observe pas ne signifie pas qu'ils sont absents, puisqu'ils ont la capacité de se déplacer et de se cacher facilement. De plus, tant pour les organismes animaux que végétaux, les espèces en phase de dispersion qui n'ont pas encore colonisé tous les habitats favorables, amènent un grand risque de fausses absences.

Une nouvelle méthode de modélisation de l'habitat potentiel, l'analyse factorielle de la niche écologique (ENFA), a été mise au point à l'IZEA (Institut de Zoologie et d'Ecologie animale) de Lausanne par Hausser, Perrin et Hirzel (Hausser, 1995 ; Hirzel & al. in prep.). Elle se base uniquement sur les données de présence de l'espèce et semble, en ce sens, être une méthode plus fiable pour l'étude des organismes animaux puisqu'on évite ainsi le risque des fausses absences.

La distribution des variables écogéographiques pour les points de présence (sites d'observation) est comparée à celle de l'ensemble des points de la zone étudiée. Cette comparaison se fait par l'intermédiaire d'un facteur de marginalité, qui indique si l'espèce occupe des milieux qui sont en marge de l'ensemble de la zone étudiée, ainsi que par d'autres facteurs limitants, les facteurs de restriction, qui permettent de voir si l'animal en question est plutôt spécialiste ou généraliste quant à ces variables. Les observations et les points de la carte sont reportés sur les axes formés par ces facteurs. On obtient ainsi, dans l'espace défini par ces nouvelles variables, un nuage de points pour les présences et un autre nuage pour tous les quadrats de la carte. Avec une telle représentation (espace en  $n$  dimensions), les points de la carte appartenant à la zone des observations seront favorables à l'espèce et une valeur élevée leur sera attribuée sur la carte de domaine potentiel, les autres points auront une valeur d'autant plus élevée que le point sera proche de la zone des points d'observation dans l'espace des variables écogéographiques.

Pour de plus amples informations, se référer à l'article de Hirzel (Hirzel et al., in prep) qui décrit cette méthode en détail. L'ENFA a déjà été utilisée pour construire les cartes de domaine potentiel et de répartition qui se trouvent dans *Mammifères de la Suisse* (Hausser, 1995).

L'avantage indiscutable de cette méthode est de ne pas nécessiter des données d'absences, cependant cela présente fatalement l'inconvénient de diminuer le nombre d'informations à disposition pour construire les cartes. Le but de ce travail est d'évaluer les avantages et inconvénients que présentent une méthode classique, le GLM, et l'ENFA. Nous avons choisi le GLM car ce dernier, couramment utilisé, permet de rendre compte d'interactions complexes. Peu de comparaisons entre les différentes méthodes de construction des cartes d'habitat potentiel ont été effectuées jusqu'à maintenant ; Manel et al. (1999), par exemple, ont comparé l'analyse discriminante, la régression logistique et le réseau neuronal artificiel pour prédire la distribution des oiseaux des rivières de l'Himalaya.

L'idéal pour une comparaison rigoureuse des différentes méthodes, serait de travailler avec une espèce dont on connaît exactement tous les paramètres, c'est-à-dire ses exigences écologiques ainsi que sa répartition exacte, hélas une telle espèce n'existe pas. Nous l'avons donc construite de toutes pièces en définissant ses exigences écologiques et la répartition qui en découle. Nous avons étudié trois cas de figure pour voir si l'une ou l'autre des analyses se trouve être plus adaptée selon les conditions données. Trois stades d'occupation du territoire ont été simulés : la colonisation depuis un territoire voisin (ici, échelle régionale), l'occupation de l'habitat optimal et la suroccupation du territoire avec débordement sur des zones moins favorables. Ces trois stades seront représentés dans ce travail par trois espèces différentes occupant la même niche écologique : *M. aliensis*, *M. helveticus* et *M. musca-cerevisiae*. Nous avons appliqué les deux types d'analyse à ces trois espèces et avons ainsi pu estimer dans quelle mesure ces analyses reflètent la réalité.

## Matériel et méthodes

### ***A. Construction du modèle***

(l'annexe I résume les étapes décrites dans cette section)

Pour générer la répartition «modèle» des trois espèces, nous avons tout d'abord construit une carte de qualité d'habitat basée sur les exigences écologiques du marsupilami. Les distributions de *M. aliensis*, *M. helveticus* et *M. musca-cerevisiae* ont ensuite été modélisées en fonction de leurs caractéristiques respectives.

### **Création de la carte de qualité d'habitat**

Le principe suivi pour créer la carte de qualité d'habitat est de préparer des cartes pour chacune des variables écogéographiques importantes pour l'écologie du marsupilami, chaque pixel contenant une valeur relative à la variable considérée. On transforme ensuite ces valeurs en « indices de qualité d'habitat ». La carte de qualité d'habitat contiendra, pour chaque pixel, une moyenne de tous ces indices, pondérée en fonction de l'importance donnée aux différentes variables écogéographiques.

#### ***1) Choix des variables***

La première étape dans la création du modèle de répartition du marsupilami a été de choisir les variables écogéographiques qui définissent son écologie et donc la qualité de son habitat. Les marsupilamis, tels que nous les avons définis, sont des espèces forestières vivant à moyenne altitude, de préférence en exposition sud. Ils aiment la proximité des points d'eau (Batem & al, 1991) et ne supportent pas les terrains trop accidentés. Bien qu'en général sensibles aux dérangements (villes, routes, chemins de fer), ces espèces apprécient la proximité des villages pour la nourriture que ceux-ci peuvent leur apporter (jardins potagers, poulaillers, etc.).

#### ***2) Préparation des cartes pour chaque variable***

Les cartes utilisées pour définir la répartition des espèces sont des cartes *rasters* (SIG) avec une résolution à l'hectare. Chaque pixel représente donc un quadrat de 100x100 m et contient la valeur représentative de cet hectare pour la variable choisie (altitude p.ex.). Les cartes contiennent un nombre considérable de données, raison pour laquelle nous avons renoncé à porter notre travail sur tout le territoire suisse. Nous avons choisi une zone de 25.5x25.5 km dans une région la plus diversifiée possible et offrant une aire propice au marsupilami suffisamment vaste. La région de Thounne a été retenue (figure a, annexe).

Dix cartes ont servi de base de travail : FORETS, ALTIT, EXPO, PENTE30, EAUX, VILLAGE, VILLE, ROUTE1, ROUTE2 et RAIL (tableau 1). Les cartes booléennes FORETS, EAUX, VILLAGE, VILLE, ROUTE1, ROUTE2 et RAIL ont été transformées en cartes de distances (tableau 1) à l'aide du logiciel *Idrisi* (Eastman, 1997). La valeur de chaque quadrat correspond alors à la distance (en mètres) qui le sépare de l'objet considéré (rivière, ville, etc.). Pour ces trois dernières cartes, supposant qu'à plus de 2 km des routes ou des voies ferrées la perturbation ne varie plus en fonction de la distance, nous avons fixé un plafond à la valeur de 2000m. Il a été fait de même pour la carte VILLE avec 5000 pour valeur maximale.

La carte EXPO (en grades) a été transformée en une carte booléenne ne contenant que les valeurs comprises entre 100 et 300, représentant les expositions sud (au sens large). Cette

dernière ainsi que FORET et PENTE30 ont été transformées avec *Biomapper* (Hirzel, 2000) en cartes de fréquences (tableau 1). Chaque pixel indique alors la fréquence d'une occurrence (forêt, pente>30° ou exposition sud), dans un rayon donné. Pour le marsupilami, ce rayon a été fixé à 500m, de telle sorte que la surface du cercle corresponde à celle de son domaine vital qui est de 0.75 km<sup>2</sup> environ.

**Tableau 1:** Les cartes de base sont des cartes dérivées de la base de donnée *Geostat* (Office Fédéral de la Statistique), recadrées sur la zone d'étude (625 km<sup>2</sup>). Dans le nom des nouvelles cartes, "D" indique qu'il s'agit d'une distance, et ".fq" d'une fréquence.

<i>Cartes de base</i>	<i>Unité</i>	<i>Contenu</i>	<i>Cartes modifiées</i>	<i>Contenu</i>
<b>Forets</b>	Booléen	Répartition des forêts	<b>Forets.fq</b>	Fréquence des forêts
			<b>Dforets</b>	Distance aux forêts (m)
<b>Altit</b>	Mètres	Altitude	<b>Altit</b>	Altitudes (m)
<b>Expo</b>	Grades	Exposition	<b>Exposud.fq</b>	Fréquence des expositions sud
<b>Pente30</b>	Booléen	Pentes de plus de 30°	<b>Pente30.fq</b>	Fréquence des pentes >30°
<b>Eaux</b>	Booléen	Réseau hydrographique	<b>Deaux</b>	Distance aux points d'eau (m)
<b>Village</b>	Booléen	Carte des villages	<b>Dvillage</b>	Distance aux villages (m)
<b>Ville</b>	Booléen	Carte de villes	<b>Dville</b>	Distance aux villes (m)
<b>Route1</b>	Booléen	Carte des routes principales	<b>Droute1</b>	Distance aux routes principales (m)
<b>Route2</b>	Booléen	Carte des routes secondaires	<b>Droute2</b>	Distance aux routes secondaires (m)
<b>Rail</b>	Booléen	Réseau ferroviaire	<b>Drail</b>	Distance aux voies ferrées (m)

### 3) Transformation des variables en « indices de qualité d'habitat »

L'étape suivante consiste à transformer ces nouvelles cartes en cartes « d'indices de qualité d'habitat », c'est-à-dire à attribuer à chaque pixel un indice (de 0 à 1) indiquant la qualité de l'habitat pour la variable considérée. Chaque variable a été transformée en fonction de l'optimum pour le marsupilami (tableau 2). Ces transformations ont été réalisées à l'aide du programme « *Marsupimapper* » (Métral & Helfer, 2000) présentant les options suivantes :

- **Variables à optimum médian :** transforme les valeurs en fonction d'une valeur optimale (moyenne, **m**) et d'une certaine tolérance (écart-type, **s**). L'algorithme se base sur la loi normale et transforme la valeur du pixel (**x**) en une probabilité selon la formule :

$$f(x, \mathbf{m}, \mathbf{s}) = \frac{1}{\mathbf{s} \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x-\mathbf{m}}{\mathbf{s}} \right)^2} \quad (\text{éq. 1})$$

Afin de ramener les valeurs entre 0 et 1, chacune d'entre elles est ensuite divisée par la probabilité maximale, donnée par  $f(x = \mathbf{m} \mid \mathbf{s})$ .

- **Variables à optimum extrême :** deux options sont encore proposées :
  1. l'indice maximum (1) est attribué aux pixels à valeur maximale (et l'indice minimum (0) aux valeurs minimales) ;
  2. l'indice maximum (1) est attribué aux pixels à valeur minimale (et l'indice minimum (0) aux valeurs maximales).

Dans les deux cas, l'algorithme repère les valeurs maximale et minimale ( $V_{max}$  et  $V_{min}$ ) de la carte et calcule l'écart entre ces valeurs ( $Ecart = V_{max} - V_{min}$ ). A la valeur de chaque pixel est soustraite  $V_{min}$ . Dans le premier cas, cette différence est divisée par l'écart entre  $V_{max}$  et  $V_{min}$ , ce qui donne l'indice ( $Indice = (x - V_{min}) / écart$ ) ; dans le deuxième cas, l'indice est donné par  $1 - ((x - V_{min}) / écart)$ .

- Le programme *Marsupimapper* permet également d'attribuer un certain poids à chaque carte, il multiplie alors simplement l'indice par le poids désiré.

**Tableau 2.** Le nom des nouvelles cartes commence par le poids attribué, la lettre "P" (probabilité) signifie qu'il ne s'agit plus de la variable proprement dite mais d'un indice de qualité d'habitat. (**m**: moyenne, **s** : écart-type ; les distributions tronquées ont un plafond à 2000 mètres pour les réseaux de transports et à 5000 mètres pour les villes.)

Cartes utilisées	Nouvelles cartes	Distribution	Paramètres	Poids
Forets.fq	6Pforets.fq	Linéaire	indice max. pour la valeur max.	6
Altit	5Paltit	Gaussienne	<b>m</b> = 900 ; <b>s</b> = 200	5
Exposud.fq	2Pexposud.fq	Linéaire	indice max. pour la valeur max.	2
Dville	2Pdville	Lin. Tronquée	indice max. pour la valeur max.	2
Dforets	1Pdforets	Linéaire	indice max. pour la valeur min.	1
Pente30.fq	1Ppente30	Linéaire	indice max. pour la valeur min.	1
Deaux	1Pdeaux	Linéaire	indice max. pour la valeur min.	1
Dvillage	1Pdville	Gaussienne	<b>m</b> = 1000 ; <b>s</b> = 600	1
Droute1	1Pdroute1	Lin. Tronquée	indice max. pour la valeur max.	1
Droute2	1Pdroute2	Lin. Tronquée	indice max. pour la valeur max.	1
Drail	1Pdrail	Lin. Tronquée	indice max. pour la valeur max.	1

#### 4) Création de la carte de qualité d'habitat

*Marsupimapper* permet enfin de grouper toutes les cartes "indice" en une carte globale pour toutes les variables. Pour cela il calcule, pour chaque pixel, la moyenne pondérée de tous les indices. Chaque valeur correspond donc au quotient de la somme des indices par le poids total. La carte globale ainsi créée est ensuite multipliée par une carte booléenne des lacs (lac=0) pour éviter d'y trouver des points de présence. L'étape finale consiste à rajouter de la stochasticité qui rend le modèle plus réaliste, simulant par exemple des animaux en déplacement ou de légères variations temporelles de la qualité de l'habitat. Cela s'est fait à l'aide d'*Idrisi* : une carte de nombres aléatoires compris entre -0.05 et 0.05 (distribution uniforme) a été additionnée à la carte globale. La carte HABITAT ainsi obtenue (figure 1) est une carte de qualité d'habitat basée sur les variables définissant l'écologie du marsupilami ; elle servira de référence tout au long de notre travail.

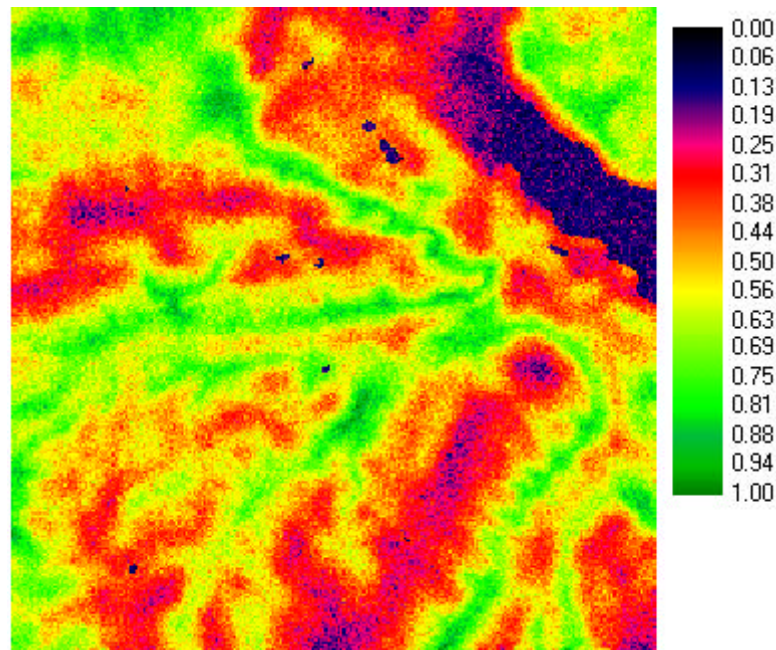


Figure 1 : Carte de qualité d'habitat du marsupilami

## Répartition des trois espèces

### 1) *M. helveticus*

*M. helveticus* est une espèce qui a colonisé la région depuis fort longtemps, les effectifs se sont stabilisés à la capacité de soutien des milieux. Elle occupe actuellement tous les endroits propices et on ne la trouve guère ailleurs.

Pour répartir cette espèce dans la région étudiée, on a tout d'abord fixé un seuil arbitraire de 0.7 à la carte de qualité d'habitat, ce qui consiste à créer une carte booléenne  $HABITAT > 0.7$ . Cette carte de répartition booléenne est basée sur un seuil extrêmement abrupte, ce qui n'a pas grand sens d'un point de vue biologique, raison pour laquelle nous avons rajouté une certaine stochasticité autour de ce seuil de la manière suivante:

1. Multiplication de la carte  $HABITAT$  par la carte booléenne  $HABITAT > 0.7$ , obtenant ainsi une nouvelle carte  $HABITAT\ HELVET$  qui présente tous les indices de qualité d'habitat plus grands que 0.7.
2. Addition d'une carte de nombres aléatoires entre  $-0.2$  et  $0.2$  (distribution uniforme) à  $HABITAT\ HELVET$ .
3. Création de la carte booléenne " $HABITAT\ HELVET > 0.7$ " que l'on appellera plutôt  $REPARTITION\ HELVET$  (figure 2a).

Avec cette méthode, on obtient une présence systématique de l'espèce dans les hectares dont l'indice est supérieur à 0.9, et les hectares avec une valeur comprise entre 0.7 et 0.9 ont d'autant plus de chance de contenir l'espèce que leur valeur est élevée, raison pour laquelle le bord des patches de la figure 2a est moins dense que le centre.



## 2) *M. musca-cerevisiae*

L'espèce *musca-cerevisiae* suit dans cette région une dynamique de population très instable. Les effectifs sont actuellement très grands et de nombreux individus vivent en dehors de leur niche idéale à cause d'une forte compétition intraspécifique.

La méthode suivie pour dessiner l'aire de répartition de *M. musca-cerevisiae* (figure 2b) est la même que celle utilisée pour *M. helveticus* mais avec un seuil de 0.5 au lieu de 0.7 (l'espèce tolère un habitat moins propice pour éviter les trop fortes densités).

## 3) *M. aliensis*

*M. aliensis* est une espèce colonisatrice dans cette région. Elle est arrivée du sud-ouest il y a quelques années ; elle a donc déjà bien colonisé les zones propices du sud-ouest de nos cartes mais sa densité diminue si l'on s'éloigne de sa zone d'entrée. Cette espèce n'a pas encore colonisé les zones propices du nord-est.

Pour simuler cette répartition, nous avons multiplié la carte de qualité d'habitat par une carte IND.DIST. contenant des indices de qualité d'habitat fonction de la distance au point d'entrée.

Ces indices ont été calculés de la façon suivante :

1. création d'une carte DISTANCE dont les valeurs représentent les distances au pixel extrême sud-ouest ;
2. transformation de ces valeurs par la fonction suivante :

$$f(dist) = \frac{1}{(dist + 150'000)^2} 100'000^2 \quad (\text{éq. 2})$$

Le principe de cette formule est de diminuer la valeur de l'indice en fonction du carré de la distance au point d'entrée. La valeur de 150'000 (mètres) est additionnée à *dist* pour simuler le fait que les individus ne rentrent pas dans le territoire par un seul pixel, ce qui ne serait pas très réaliste, le foyer source est ainsi en dehors de la carte ; cette nouvelle distance est ensuite multiplié par 100'000<sup>2</sup>, ce facteur diminue l'effet de la distance sur la fonction (en effet, une fonction directement inverse au carré de la distance impliquerait une diminution trop rapide des valeurs par rapport à la distance, pour la répartition que l'on désire) ;

3. division des valeurs ainsi obtenues par la valeur maximale pour obtenir des indices allant de 0 à 1.

De même que pour *M. helveticus*, un seuil de 0.7 a été fixé et de la stochasticité a été ajoutée. On obtient ainsi la carte de distribution de l'espèce *M. aliensis* (figure 2c).

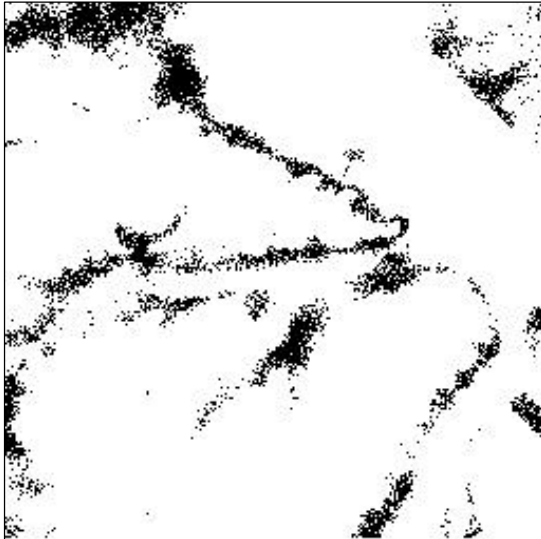


Figure 2a

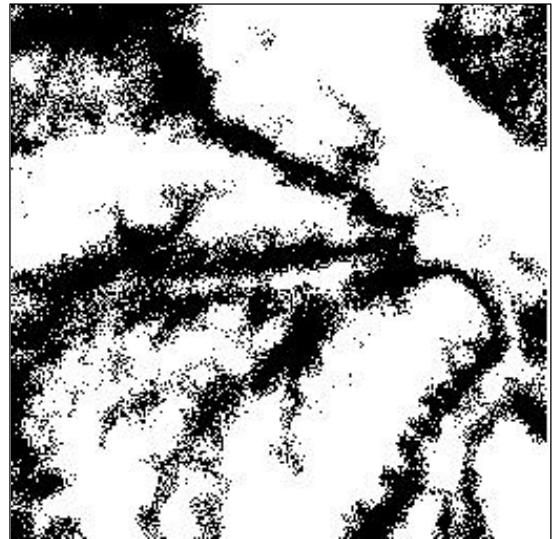


Figure 2b

**Figure 2 :** Répartition des trois espèces.  
Ci-dessus : a) *M. helveticus*, b) *M. muscivorensis*.  
Ci-contre : c) *M. aliensis*.

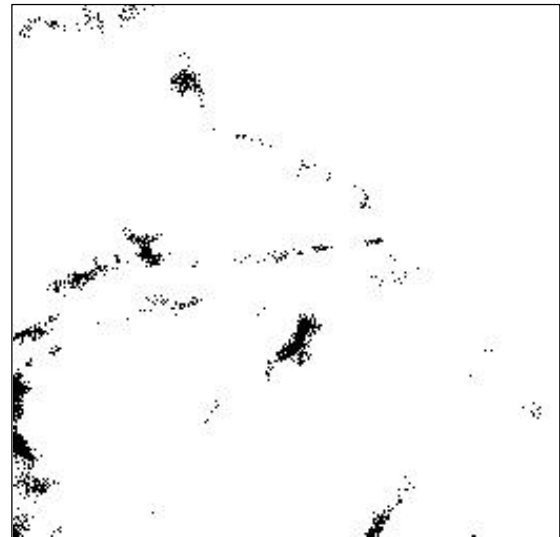


Figure 2c

## ***B. Analyses de la niche***

Les étapes décrites dans la section A ont servi à construire le modèle « marsupilami », les données qui y figurent ne sont pas accessibles dans la réalité. Les étapes qui suivent correspondent aux étapes qui s'effectuent habituellement pour des espèces réelles. Nous avons simulé tout d'abord un échantillonnage « sur le terrain » puis appliqué les deux analyses (ENFA et GLM) aux données ainsi récoltées.

### **Observations sur le terrain**

#### ***Points de présence***

Dans la réalité, les seules données auxquelles on a accès sont des données d'observation. Nous avons simulé un échantillonnage aléatoire simple sur l'ensemble de la zone étudiée

pour les obtenir. Pour ce faire, nous avons échantillonné à l'aide du module *Sampler* (*Biomapper*) le 10% des points constituant la carte de répartition des espèces *M.helveticus* et *M. musca-cerevisiae* (figures 2a et 2b). On garde ainsi le même effort d'échantillonnage. Pour *M. aliensis*, le 40% des points a été retenu car l'espèce est plus rare (figure 2c). Cet écart quant à l'effort d'échantillonnage est justifié; en effet, si l'on sait que l'espèce étudiée est une espèce rare ou en phase de colonisation, on essaiera d'avoir des pièges suffisamment efficaces et nombreux pour assurer un bon prélèvement.

### **Points d'absence**

Avant d'échantillonner les points d'absence, nous avons transformé (option *Buffer* dans *Idrisi*) les cartes de présence de manière à ce que tous les quadrats situés à moins de 700m d'un point de présence soient considérés comme des sites occupés par l'espèce. Nous tenons ainsi compte de l'auto-corrélation spatiale et réduisons grandement le risque de fausses absences. Les points d'absence ont ensuite été obtenus par un échantillonnage aléatoire sur le négatif de la carte de présence « bufferisée ». Les pixels ont été tirés de façon à obtenir le même nombre de données d'absences que de présences.

## **L'analyse factorielle de la niche écologique (ENFA)**

L'ensemble de cette analyse a été effectuée à l'aide de *Biomapper*.

Nous avons basé l'analyse sur 12 cartes de variables écogéographiques, dont 6 ont déjà servi à définir l'habitat de l'espèce (section A). Les cartes retenues sont : ALTIT, EXPOSUD.FQ, DVILLE, DFORET, DROUTE1, PENTE30.FQ. Les nouvelles variables introduites sont la distance aux rivières, la distance aux lacs, la distance aux pâturages, la distance aux prés, la fréquence des rochers et celle des broussailles. Les cartes (tableau 3) ont été construites selon le même procédé qu'énoncé plus haut (section A, §2).

Cette façon de procéder permet de rendre les conditions d'analyse assez semblables à ce qui se fait habituellement. En effet, lors d'un travail de terrain, on n'est jamais sûr de relever toutes les variables écologiques déterminantes pour l'espèce en question.

**Tableau 3** : Les cartes de base sont des cartes dérivées de la base de donnée *Geostat* (Office Fédéral de la Statistique), recadrées sur la zone d'étude (625 km<sup>2</sup>). Dans le nom des nouvelles cartes, "D" indique qu'il s'agit d'une distance, et ".fq" d'une fréquence.

<i>Carte de base</i>	<i>Unité</i>	<i>Contenu</i>	<i>Cartes modifiées</i>	<i>Contenu</i>
<b>Riv</b>	Booléen	Carte des rivières	<b>Driv</b>	Distance (m) aux rivières
<b>Lac</b>	Booléen	Carte des lacs	<b>Dlacs</b>	Distance (m) aux lacs
<b>Patur</b>	Booléen	Carte des pâturages	<b>Dpatur</b>	Distance (m) aux pâturages
<b>Pres</b>	Booléen	Carte des prés	<b>Dpres</b>	Distance (m) aux prés
<b>Roches</b>	Booléen	Carte des roches	<b>Roches.fq</b>	Fréquence des roches
<b>Brous</b>	Booléen	Carte des broussailles	<b>Brous.fq</b>	Fréquence des broussailles

Les cartes de distance aux rivières et aux lacs sont corrélées à la carte de distance aux eaux utilisée pour définir l'espèce.

Les conditions d'utilisation de l'ENFA exigent des variables normales (Hirzel & al, in prep.). Après avoir introduit les cartes choisies dans *Biomapper*, nous les avons transformées selon la méthode Box-Cox (Sokal et Rohlf, 1981). Cette option trouve la meilleure fonction pour normaliser chaque variable.

Les points de présence sont répartis aléatoirement en deux lots de taille égale, à l'aide du module *Sampler*. On obtient alors une carte de calibration (PRCALIB.BL) pour construire le modèle et une carte de validation (PRVALID.BL) qui permettra de le tester sur des données indépendantes. Les cartes ainsi obtenues se trouvent en annexe (figures b, c, d, e, f, g).

*Biomapper* centre et réduit les variables avant d'appliquer l'ENFA qui calcule une matrice de covariance pour les points de présence et en extrait les valeurs propres (variance expliquée par les nouveaux facteurs) et les scores des variables sur les nouveaux axes définis (facteurs). Le premier facteur correspond à la marginalité, les suivants sont des facteurs limitants. Pour l'espèce une marginalité et une restriction globales sont également fournies par cette analyse.

Pour calculer la carte de qualité de l'habitat, nous avons retenu le nombre de facteurs considérés comme significatifs par comparaison à la distribution du Broken-Stick et qui explique au moins le 80% de la variance du modèle. Le nombre de catégories a été fixé à six pour les trois espèces. De ce nombre de catégories dépendra la finesse de la carte.

## Le modèle linéaire généralisé (GLM)

Le modèle linéaire généralisé a été construit à l'aide du logiciel *S-Plus* (-, 1999). Les variables écogéographiques utilisées sont les mêmes que pour l'ENFA. En ce qui concerne les données d'observations, elles contiennent les mêmes données de présence que l'ENFA (PRESCALIB.BL) ainsi que les données d'absences (ABSCALIB.BL). Nous avons construit une carte qui contient ces deux informations à la fois à l'aide d'*Idrisi*.

On a effectué l'analyse GLM pas à pas\* (option *stepwise*) sur *S-Plus*. Cette analyse consiste à ajouter et enlever les variables au modèle l'une après l'autre pour voir si elles sont significatives. Elle a tendance à retenir les premières variables introduites, ce qui influence le résultat de l'analyse. Nous avons donc choisi d'introduire les cartes selon leur importance décroissante pour l'écologie du marsupilami telle que nous l'avons fixée, ce qui donne au GLM toutes les chances de retenir les variables les plus pertinentes. Les interactions n'étant pas toujours linéaires, le terme simple a été introduit suivi du terme quadratique.

Une fois l'option *stepwise* terminée, une Anova (test du  $c^2$ ) permet de tester la réduction de déviance du système amenée par chacune des variables retenues. Les variables non significatives (seuil : 0.05) sont enlevées du modèle et l'option *stepwise* est alors relancée, et ainsi de suite jusqu'à ce que toutes les variables soient significatives. Un test de *t* permet de s'assurer que les coefficients, poids de chaque variable dans le modèle, sont significativement différents de 0. S'ils ne sont pas significatifs il faut encore enlever la variable en question.

\* Les commandes utilisées dans *S-Plus* sont les suivantes (les commandes sont en caractères gras, les paramètres en italique) :

`nom_step(glm(var réponse ~ var descr1 + var descr2 etc., data = fichier de données brutes, family = binomial), direction = "both", trace = T)`

Le modèle obtenu nous permet de calculer une valeur pour chaque pixel. Les valeurs sont ensuite transformées\* (option *Stretch* d'*Idrisi*) de telle sorte qu'elles soient comprises entre 0 et 100, la carte représentera alors la qualité d'habitat de l'espèce.

Avant la validation, les cartes de domaine potentiel ont été lissées au moyen d'un filtre gaussien. Cette opération a pour but de lisser la distribution des indices de qualité d'habitat.

En effet, en raison du nombre faible de catégories retenues dans l'ENFA, cette distribution est en escaliers. Le lissage consiste à attribuer à un pixel une valeur dépendante des pixels qui l'entourent. Dans le cas présent, le pixel central d'une aire de 7x7 pixels prend une valeur moyenne pondérée des pixels alentours, chaque pixel contribuant par un poids dépendant de sa distance au point central selon une distribution normale. L'aire de 7x7 pixels a été retenue après avoir validé par un indice de corrélation (cf. validation des modèles) des cartes obtenues avec différentes aires de lissage. L'aire de 7x7 semble l'aire optimale. Une courbe de l'effet du lissage sur le coefficient de corrélation (cf. validation des modèles, section C) justifiant le seuil choisi se trouve en annexe (figure h).

Le lissage a aussi été effectué sur les cartes obtenues avec le GLM afin d'assurer le même traitement aux deux analyses (le lissage n'est pas si nécessaire pour le GLM, la distribution étant déjà lisse; dans certains cas, elle améliore néanmoins le modèle). Nous avons retenus, entre les cartes lissées ou originales, les cartes qui présente la meilleure corrélation avec la carte de référence (HABITAT).

### ***C. Validation des modèles***

Les cartes de domaine potentiel obtenues par les deux analyses ont été validées de façon identique. Les différents tests de validation ont été effectués sur *Biomapper* à l'aide du module de *validation*.

#### ***Comparaisons avec la carte d'habitat de référence (HABITAT)***

Pour chacune des méthodes, un test d'évaluation comparant la carte de domaine potentiel obtenue par le modèle à la carte de qualité d'habitat réelle (figure 1) a été effectué. Dans la réalité, cette comparaison n'est pas possible car cette deuxième carte fait défaut. Nous avons calculé le coefficient de corrélation entre ces deux cartes pour les trois espèces selon les deux analyses. Pour éviter une auto-corrélation spatiale (Bonham-Carter, 1994), nous n'avons pris qu'une partie des points de chaque carte pour ce calcul. Le nombre de points tirés au sort a été fixé à 1% du nombre total.

#### ***Evaluation des modèles par les cartes de validation***

La validation des modèles se fait, comme pour les espèces réelles, à partir du deuxième lot de points d'observations, c'est-à-dire les cartes de validation.

\* Le procédé est identique à celui utilisé par *Marsupimapper* pour les optima extrêmes (section A, §3, point 1) et les variables obtenues sont ensuite multipliées par 100.

La première évaluation, proposée par Hirzel (Hirzel, 2000), consiste à calculer la proportion de points de présence ( $X$ ) situés dans des pixels qui ont une valeur supérieure à 50 sur la carte de domaine potentiel produite, ainsi que la proportion de points totaux ( $Y$ ) de la carte de domaine potentiel supérieurs à 50. Pour une même espèce, le modèle sera d'autant plus fiable que  $X$  est élevé et que l'écart entre  $X$  et  $Y$  est grand.

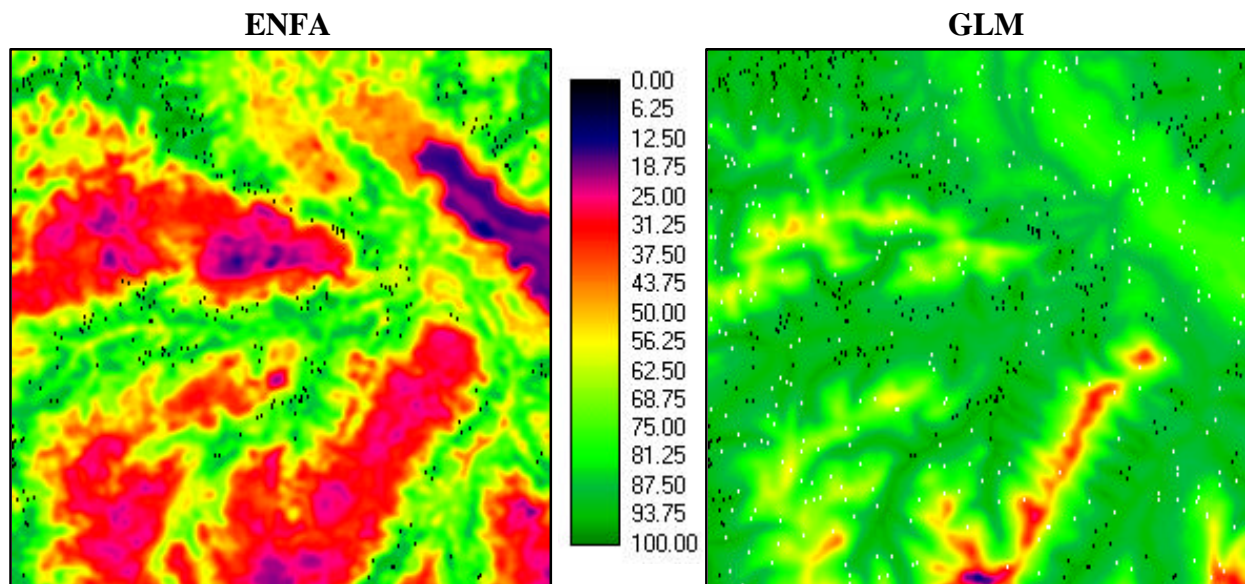
Le deuxième test, la technique ROC (Fielding & Bell, 1997), est une statistique globale permettant d'évaluer à quel point le modèle obtenu est différent d'un modèle aléatoire

( $AUC = 0.5$ ). L'aire sous la fonction ( $AUC$ , area under the curve) ainsi obtenue donne une bonne évaluation de l'adéquation globale du modèle.

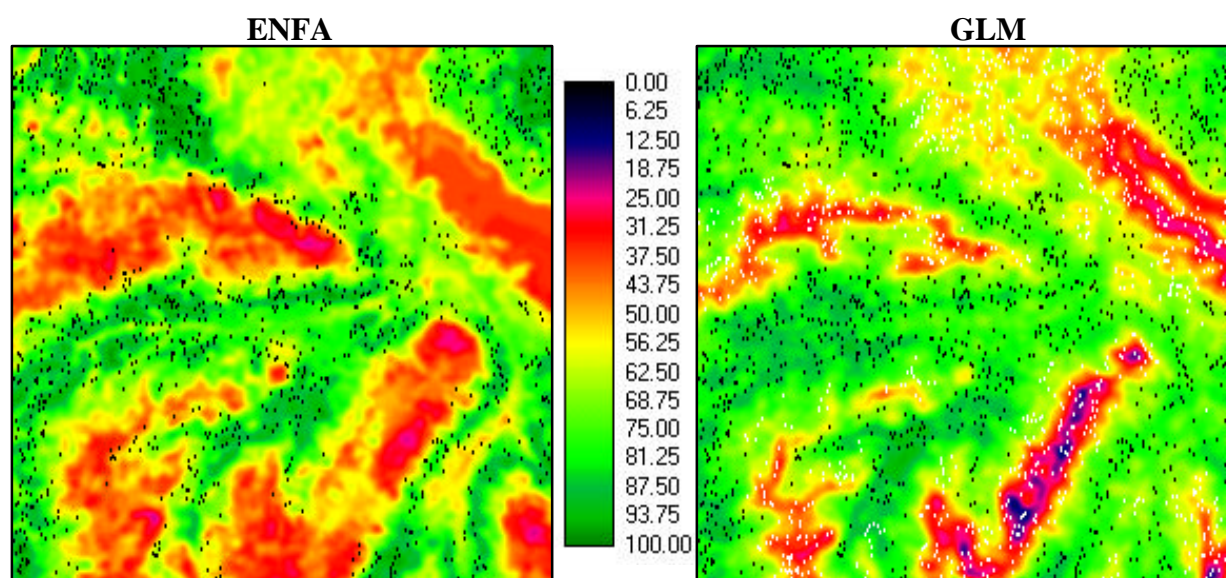
## Résultats

### Cartes obtenues avec l'ENFA et le GLM

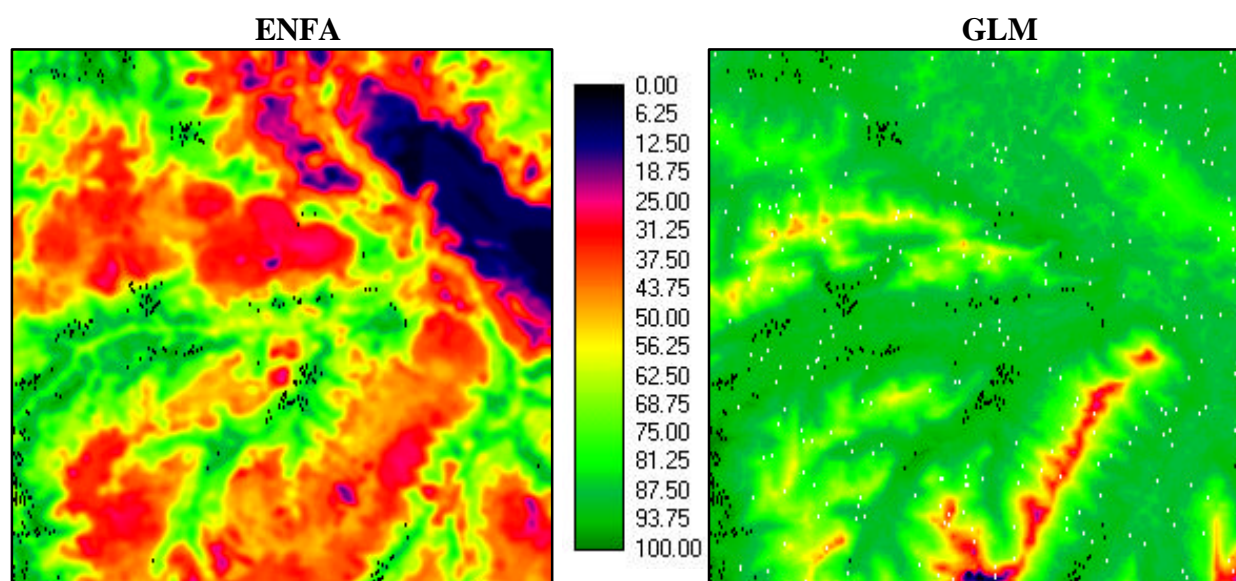
Les cartes obtenues avec l'analyse factorielle de la niche écologique et le modèle linéaire généralisé sont présentées dans les figures ci-dessous. Toutes ces cartes ont été lissées, à l'exception de celle obtenue par le GLM pour *M. aliensis*. Les points d'observation de présence ont été superposés à ces cartes, les points d'absence n'ont été signalés que pour le GLM puisque l'ENFA n'en tient pas compte.



**Figure 5 :** Cartes de domaine potentiel de *M. helveticus* selon l'ENFA (à gauche) et le GLM (à droite). Les points représentent les points utilisés pour la calibration des modèles, présences en noir (n=305), absences en blanc (n=319).



**Figure 6 :** Cartes de domaine potentiel de *M. musca-cerevisiae* selon l'ENFA (à gauche) et le GLM (à droite). Les points représentent les points utilisés pour la calibration des modèles, présences en noir (n=1100), absences en blanc (n=1063).



**Figure 7 :** Cartes de domaine potentiel de *M. aliensis* selon l'ENFA (à gauche) et le GLM (à droite). Les points représentent les points utilisés pour la calibration des modèles, présences en noir (n=291) absences en blanc (n=301).

Les cartes basées sur l'ENFA sont contrastées et varient peu d'une espèce à l'autre. Avec le GLM, la carte de l'espèce *musca-cerevisiae* se distingue de celles des deux autres espèces, ces dernières étant relativement uniformes.

## Variables retenues par les analyses

### *L'analyse factorielle de la niche écologique*

Pour les trois espèces, l'ensemble des scores obtenus pour les variables sur les différents facteurs se trouvent en annexe. Le tableau 4 indique différents paramètres concernant l'ENFA. La variance expliquée dépend entre autre du nombre de facteurs retenus.

**Tableau 4 :** Paramètres de l'ENFA pour les trois espèces.  
Le nombre de facteurs retenus a été choisi par comparaison avec la distribution du Broken-Stick.

	<i>M. helveticus</i>	<i>M. musca-cerevisiae</i>	<i>M. aliensis</i>
Nb. de facteurs retenus	9	7	5
Variance expliquée	94.6%	79.9%	85.4%
Marginalité globale	0.610	0.527	0.692
Tolérance globale	0.569	0.754	0.490

Les trois variables expliquant le mieux la marginalité de *M. helveticus* sont les distances aux forêts, aux rivières et l'altitude (tableau 5) ; la distance aux villes n'a que peu d'influence. Le premier facteur de restriction indique que cette espèce est particulièrement sensible à la distance aux forêts et à l'altitude (tableau 5), la distance aux routes principales étant moins importante.

Pour *M. musca-cerevisiae*, la distance aux forêts a une grande influence sur la marginalité (tableau 5); l'exposition et la pente jouent un moindre rôle. Le premier facteur de restriction indique une spécialisation en fonction de l'altitude et de la distance aux lacs (tableau 5) ; la fréquence des pentes supérieures à 30° et la distance aux routes principales ont peu d'importance.

En ce qui concerne *M. aliensis*, les trois variables expliquant le mieux la marginalité sont les distances aux forêts, aux rivières et aux villes (tableau 5); l'exposition et la pente n'ont que peu d'influence. Le premier facteur de restriction indique qu'elle est particulièrement sensible à l'altitude, aux distances aux lacs et aux forêts (tableau 5) ; la fréquence des pentes supérieures à 30° par contre ne restreint que peu sa niche.

**Tableau 5 :** Variables ordonnées selon leur score sur les deux premiers facteurs.

Pour le facteur 1 (marginalité), le signe des valeurs indique si la moyenne de l'espèce est inférieure ou supérieure à la moyenne de tous les pixels. Pour le facteur 2 (1<sup>er</sup> facteur de restriction) les valeurs sont à interpréter en valeurs absolues. Les variables auxquelles un poids a été attribué pour la distribution de l'espèce sont indiquées en caractère gras.

HELVETICUS		MUSCA-CEREVISIAE		ALIENSIS	
Facteur 1	Facteur 2	Facteur 1	Facteur 2	Facteur 1	Facteur 2
<b>DVILLE (0.04)</b>	DRROUTE1 (0.01)	PENTE30 (-0.05)	DRROUTE1 (0.02)	<b>EXPO (0.04)</b>	PENTE30 (0.05)
<b>EXPO (-0.10)</b>	<b>DVILLE (-0.03)</b>	<b>EXPO (-0.07)</b>	PENTE30 (0.03)	PENTE30 (-0.05)	BROUS.FQ (0.09)
PENTE30 (-0.11)	PENTE30 (-0.03)	DRROUTE1 (-0.11)	<b>DVILLE (0.11)</b>	DRROUTE1 (-0.20)	DPATUR (-0.09)
DRROUTE1 (-0.14)	DRIV (0.04)	<b>DVILLE (0.18)</b>	BROUS.FQ (0.11)	BROUS.FQ (-0.23)	DPRES (-0.09)
DPRES (-0.23)	<b>EXPO (0.04)</b>	<b>ALTIT (-0.24)</b>	DRIV (-0.14)	<b>ALTIT (-0.24)</b>	DRIV (-0.10)
BROUS.FQ (-0.26)	ROCHES.FQ (0.05)	DPATUR (0.25)	DPRES (-0.15)	DPRES (-0.24)	<b>EXPO (-0.11)</b>
DLACS (0.32)	BROUS.FQ (-0.06)	BROUS.FQ (-0.28)	<b>EXPO (-0.17)</b>	DPATUR (0.24)	ROCHES.FQ (-0.11)
ROCHES.FQ (-0.32)	DPATUR (0.06)	DPRES (-0.28)	ROCHES.FQ (0.21)	ROCHES.FQ (-0.27)	DRROUTE1 (0.12)
DPATUR (0.32)	DPRES (0.08)	DRIV (-0.32)	DPATUR (-0.25)	DLACS (0.28)	<b>DVILLE (0.15)</b>
<b>ALTIT (-0.33)</b>	DLACS (-0.09)	DLACS (0.34)	<b>DFORETS (-0.28)</b>	<b>DFORETS (-0.42)</b>	<b>DFORETS (-0.23)</b>
DRIV (-0.38)	<b>ALTIT (-0.13)</b>	ROCHES.FQ (-0.38)	DLACS (0.36)	DRIV (-0.44)	DLACS (0.25)
<b>DFORETS (-0.53)</b>	<b>DFORETS (0.18)</b>	<b>DFORETS (-0.55)</b>	<b>ALTIT (0.51)</b>	<b>DVILLE (0.45)</b>	<b>ALTIT (0.34)</b>

### Le modèle linéaire généralisé

Les variables retenues pour l'espèce *M. helveticus* ainsi que leur coefficient se trouvent dans le tableau 6. Le modèle n'a retenu que trois termes dont un terme au carré, ainsi que l'ordonnée à l'origine. Seules deux variables écogéographiques ont été retenues : l'altitude et la distance aux rivières. Le carré de la distance aux rivières a le coefficient le plus élevé. L'ordonnée à l'origine a une valeur négative.



Pour l'espèce *M. musca-cerevisiae* (tableau 7), douze termes ont été retenus, ainsi que l'ordonnée à l'origine. Cela représente huit variables écogéographiques. La distance aux rivières est le terme qui a le coefficient le plus élevé. Il est suivi de la distance aux pâturages, puis de la distance aux lacs. L'ordonnée à l'origine est proche de zéro.

Dix termes ont été retenus pour le modèle de *M. aliensis* (tableau 8), ainsi que l'ordonnée à l'origine qui a une valeur très négative. Parmi les sept variables écogéographiques retenues, c'est la distance aux pâturages qui a le coefficient le plus élevé, suivi par l'altitude.

**Tableau 6 :** Modèle obtenu par le GLM pour *M. helveticus*. ( $R^2 = 0.333$ , degrees of freedom : 604 total, 600 residual. Null Deviance : 150.9851, Residual Deviance: 100.0512)

Ordonnée à l'origine	-0.4076391
Altit	0.002428518
Altit <sup>2</sup>	-1.131691e-006
Driv <sup>2</sup>	-0.003779957

**Tableau 7 :** Modèle obtenu par le GLM pour *M. musca-cerevisiae*. ( $R^2 = 0.608$ , degrees of freedom: 2146 total, 2133 residual. Null Deviance : 536.1603, Residual Deviance : 208.906)

Ordonnée à l'origine	-0.6475475
Altit	0.001738869
Altit <sup>2</sup>	-6.905645e-007
Expsud.fq <sup>2</sup>	-4.990586e-008
Dville	0.00004103656
Dville <sup>2</sup>	-1.546773e-009
Dforets	-0.001257898
Dforets <sup>2</sup>	1.070628e-006
Driv	0.02907169
Driv <sup>2</sup>	-0.003866966
Dlacs	0.01429541
Dpatur	0.02291386
Dpres <sup>2</sup>	-0.0009099736

**Tableau 8 :** Modèle obtenu par le GLM pour *M. aliensis* ( $R^2 = 0.719$ , degrees of freedom: 586 total, 575 residual. Null Deviance: 812.3412, Residual Deviance: 223.7512)

Ordonnée à l'origine	-32.40603
Altit	0.06416776
Altit <sup>2</sup>	-0.00003493465
Dville	0.0003804302
Dforets	-0.009857871
Droute1	0.03262162
Droute1 <sup>2</sup>	0.001913297
Pente30	-2.125056e-007
Driv <sup>2</sup>	-0.0235586
Dpatur	-0.6400716
Dpatur <sup>2</sup>	0.07509666

## Validation des modèles

Pour la comparaison des cartes de domaine potentiel à la carte de référence HABITAT (figure 1), les indices de corrélation (tableau 9) sont tous significativement différents entre les deux types d'analyse. Le coefficient de corrélation est plus élevé avec l'ENFA pour HELVETICUS et ALIENSIS alors que pour MUSCA-CEREVISIAE la valeur obtenue est plus grande avec le GLM.

**Tableau 9 :** Coefficients de la corrélation ( $R^2$ ) entre les cartes de domaine potentiel et la carte HABITAT et comparaison des  $R^2$  par un test de Student. DL : HELV. : 644, 656 ; MUSCA : 678, 676 ; ALI. : 703, 681. (\*\*hautement significatif, \*\*\*très hautement significatif, seuil : 0.05)

Cartes	HELVETICUS		MUSCA-CEREVISIAE		ALIENSIS	
	ENFA	GLM	ENFA	GLM	ENFA	GLM
$R^2$	0.515	0.392	0.585	0.670	0.546	0.341
$P$ (student)	0.001**		0.004**		<0.001***	

Les résultats des tests de validation courants (tableau 10), basés sur les données d'échantillonnage, montrent (validation 1) que les cartes issues du GLM sont peu contrastées. Dans les trois cas, plus du 90% de leurs pixels ont une valeur supérieure à 50, ce qui n'est pas le cas avec l'ENFA. Pour HELVETICUS et ALIENSIS, l'indice de ROC est plus élevé avec l'ENFA alors que pour MUSCA-CEREVISIAE il est plus grand avec le GLM.

**Tableau 10 :** Résultats des validations pour les deux analyses. ( $Auc$ =area under the curve)

Valid.	Cartes Analyses	HELVETICUS		MUSCA-CEREVISIAE		ALIENSIS	
		ENFA	GLM	ENFA	GLM	ENFA	GLM
1	Prop.de pts. prés.>50	0.981	1.000	0.983	1.000	0.988	1.000
	Prop. de pixels>50	0.592	0.989	0.755	0.808	0.493	0.978
2	$Auc$ (ROC)	0.844	0.902	0.748	0.980	0.924	0.937

## Discussion

Le but de notre travail était de comparer l'ENFA au GLM. Pour cela, nous avons systématiquement utilisé le même nombre de points de présence d'une analyse à l'autre. Il faut savoir cependant que l'on n'avait pas la même quantité de points d'observation pour toutes les espèces. Avec un même effort d'échantillonnage, nous obtenons plus de point de présence pour l'espèce *M. musca-cerevisiae* que pour l'espèce *M. helveticus*. Ceci est cohérent par rapport à ce qu'on aurait obtenu si on avait effectué un véritable travail de terrain. En effet, l'espèce *M. musca-cerevisiae* étant en surpopulation, on l'observera plus fréquemment. L'espèce *M. aliensis*, quant à elle, fait l'objet d'un effort de piégeage plus important pour palier au manque de données dû à la rareté de l'espèce. Ces différences de quantité d'observations empêchent toute comparaison rigoureuse entre espèces.

Les deux analyses ont globalement bien reproduit la répartition des habitats favorables. On retrouve les mêmes patterns d'une analyse à l'autre, mais le contraste entre les différentes zones diffère. L'analyse factorielle de la niche écologique aboutit à des cartes plus contrastées que celles obtenues par le modèle linéaire généralisé, à l'exception de MUSCA-CEREVISIAE, ce qui les rend plus ajustées à la carte de référence (HABITAT). Ce phénomène qui s'observe déjà visuellement sur les cartes, est confirmé par l'interprétation des résultats des différents tests. Si tous les points de présence (cf. validation 1) sont plus grands que 50 pour le GLM, ce qui n'est pas le cas de l'ENFA, la proportion de tous les pixels de la carte plus grands que 50 pour HELVETICUS et ALIENSIS est beaucoup trop élevée pour que ces modèles soient bons. On sait en effet que la majorité des hectares n'est pas optimale pour le marsupilami (figure 1). Autrement dit, les modèles obtenus avec le GLM sont trop permissifs et semblent surestimer la qualité de l'habitat, du moins pour *M. aliensis* et *M. helveticus*, ce qui est confirmé par le coefficient de corrélation qui compare

les cartes à la carte de référence. Pour *M. musca-cerevisiae*, ce coefficient montre que le GLM est meilleur, ce qui tendrait à dire que dans ce cas, l'ENFA est trop restrictive.

Deux explications peuvent être formulées quant à la différence de performance du GLM dans le cas de *M. musca-cerevisiae*. Premièrement, le nombre de données d'observation pour les deux analyses est plus élevé dans ce cas que pour les deux autres espèces, laissant supposer que le GLM est d'autant plus efficace que les données sont nombreuses. L'ENFA quant à elle donne des résultats semblables, indépendamment du nombre de données et de l'espèce. L'explication alternative serait que c'est le seuil de tolérance (0.5) donné à *M. musca-cerevisiae* pour simuler le débordement de sa niche qui rend le GLM plus performant.

Cette seconde hypothèse semble toutefois infondée : on constate une tendance de la GLM à surestimer la qualité de l'habitat, tant pour *M. helveticus* que pour *M. aliensis* qui ont un seuil de 0.7 ; il serait alors surprenant que le GLM devienne plus restrictif pour une espèce moins exigeante (seuil de 0.5).

La première hypothèse semble par contre très plausible, la densité de points d'observation (absences et présences) utilisés pour calibrer le modèle de cette espèce étant particulièrement grande (elle correspond environ à 3 observations par km<sup>2</sup>). Il est fort probable que la répartition des observations qui en résulte soit également un facteur important pour la qualité du modèle. Avec beaucoup de données, les points d'absences sont structurés et n'ont plus le caractère aléatoire que l'on observe dans les autres cas. Cela est dû à la « bufferisation » effectuée afin de tenir compte de l'auto-corrélation spatiale (cf. section B, § Points d'absence). Ainsi, avec beaucoup de données, on obtient des patches de présence dans lesquels les absences sont impossibles, ce qui limite les risques de fausse absence. Nous disposons de plus d'informations valides qu'avec l'ENFA, ce qui expliquerait que le GLM est meilleur dans ce cas.

Les deux tests de fiabilité en rapport aux points d'observations (présences - absences) qui ont permis la construction de la carte de domaine potentiel, donnent des résultats contradictoires. Le ROC indique que le GLM est meilleur dans les trois cas étudiés alors que la première validation indique le contraire. L'indice de corrélation, qui permet de voir comment les cartes de domaine potentiel s'ajustent à la carte de qualité de l'habitat, ainsi que l'analyse visuelle des cartes indiquent quant à eux que l'ENFA est meilleure pour *M. helveticus* et *M. aliensis* et moins performante dans le cas de *M. musca-cerevisiae*. Ces résultats rendent attentifs au fait que les résultats de la méthode ROC sont à interpréter avec précaution. Cette méthode indique la qualité du modèle par rapport aux données utilisées mais il serait abusif d'extrapoler ses résultats aux exigences écologiques de l'espèce.

En effet, la méthode ROC, testant la qualité du modèle avec différents seuils, semble efficace pour déterminer si les zones délimitées en terme de qualité d'habitat suivent la répartition réelle des zones favorables à l'espèce, mais ne parvient pas à apprécier le niveau d'exigence de cette espèce ; elle ne détecte pas le manque de contraste entre les différentes zones. Cette méthode ne semble donc pas appropriée à l'évaluation de cartes d'habitat définies par des indices de qualité mais serait plus adéquate pour la validation de cartes booléennes de répartition.

Pour créer notre modèle, nous avons donné un poids important à l'altitude, l'exposition, la distance aux villes ainsi qu'au caractère forestier de l'espèce. L'ENFA a généralement bien tenu compte de ces paramètres, de façon relativement semblable pour les trois espèces (tableau 5). Le GLM, s'il a généralement retenu ces paramètres, a souvent donné une grande importance à des variables dont on n'avait pourtant pas tenu compte en créant le marsupilami (tableau 6). C'est le cas notamment pour l'espèce *M. musca-cerevisiae* qui a

donné un bon résultat avec cette analyse. Cela provient sans doute du nombre de variables écogéographiques retenues pour construire le modèle.

Il est intéressant de noter que selon les deux analyses, les trois espèces semblent éviter les lacs (tableaux 5 et 7) alors que nous avons établi que le marsupilami appréciait la proximité de l'eau. Ceci est dû au fait que les lacs se situent généralement dans des zones défavorables, à cause de la proximité des villes ou de l'altitude par exemple (la majorité des villes se situent autour du lac de Thoun). Nous avons donné un poids de 2 aux villes et de 5 à l'altitude, c'est pourquoi l'effet de ces variables est plus fort que celui de la distance aux lacs. Il faut donc tenir compte des interactions qui peuvent exister entre les différentes variables et ne pas interpréter tous les résultats en terme de comportement.

La tolérance globale obtenue par l'ENFA varie d'une espèce à l'autre. La différence entre *M. helveticus* et *M. musca-cerevisiae* est facilement compréhensible puisque le seuil de tolérance fixé pour construire la répartition de *M. musca-cerevisiae* est plus bas (0.5 contre 0.7 pour *M. helveticus*). Le cas de *M. aliensis* est plus complexe. Selon l'ENFA, sa tolérance est particulièrement faible alors que le seuil fixé est identique à celui de *M. helveticus*. Il est probable que cela soit dû au point d'entrée choisi. En effet, ce point est à l'opposé des villes et des routes principales, et la distance à ces objets a été retenue comme déterminante dans l'écologie de l'espèce. Ces variables sont en effet plus corrélées au facteur de restriction pour cette espèce que pour les autres (tableau 5).

Il faut savoir que les tolérances et marginalités globales sont relatives à la zone d'étude (Hirzel & al., in prep); on constate ici qu'elles dépendent également du niveau de saturation de la niche. Elles ne sont donc pas à interpréter comme des caractéristiques absolues de l'espèce.

Au vu des résultats obtenus pour nos trois espèces virtuelles, il semble que les performances du GLM sont le résultat d'un compromis entre la quantité d'informations et la qualité de ces informations. Plus le nombre d'observations est grand, mieux elles sont structurées et le risque d'erreur s'en trouve diminué. La plus grande quantité d'informations dont dispose le GLM par rapport à l'ENFA n'est utile que si le risque que cette information supplémentaire soit erronée est faible. Dans le cas d'une espèce en phase de colonisation, comme *M. aliensis*, les fausses absences sont nombreuses. Au contraire, pour une espèce qui déborde de sa niche comme *M. musca-cerevisiae*, les fausses absences sont peu nombreuses, augmentant ainsi la qualité de l'information. Pour le cas intermédiaire d'une espèce stable, notre comparaison plaide en faveur de l'ENFA.

L'ENFA semble donc plus robuste que le GLM quand à la qualité et la quantité de données. Il serait intéressant d'approfondir ces résultats en testant l'influence du nombre de données en lui-même sur ces analyses.

L'utilisation d'une espèce virtuelle nous a permis d'évaluer avec assurance les méthodes de construction de cartes d'habitat puisque l'on connaît parfaitement ses caractéristiques. En l'occurrence, elle nous rend également attentifs aux précautions à prendre dans l'interprétation des variables retenues par les analyses. Ces variables sont utilisables dans la construction de cartes de domaine potentiel mais la corrélation observée entre la présence de l'animal et les valeurs pour chaque variable ne doit pas être interprétée comme une relation de cause à effet dans une approche comportementale.

Le marsupilami est un outil très précieux qui pourrait être fort utile à de nombreuses autres analyses. Nous avons remarqué que les méthodes de validation des modèles proposées par *Biomapper* (tableau 9) aboutissent à des résultats parfois divergents. Il serait intéressant d'utiliser le concept d'espèce virtuelle pour évaluer jusqu'à quel niveau leurs résultats peuvent être interprétés.

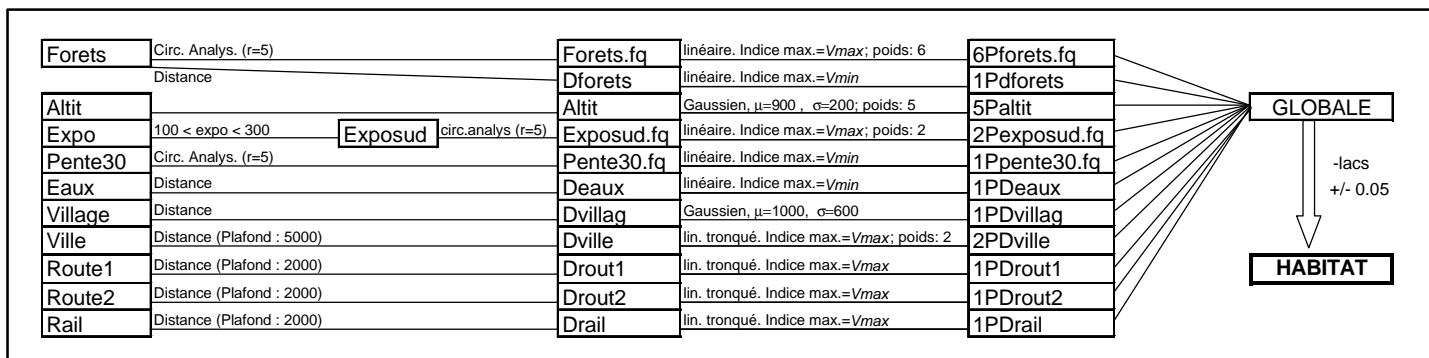
## Remerciements

Nous tenons à remercier Alexandre Hirzel pour sa patience et la disponibilité dont il a fait preuve tout au long de ce travail. Par ses précieux conseils et son humour, il a rendu notre initiation aux SIG et à la programmation moins douloureuse et même, n'ayons pas peur des mots, presque agréable.

## Bibliographie

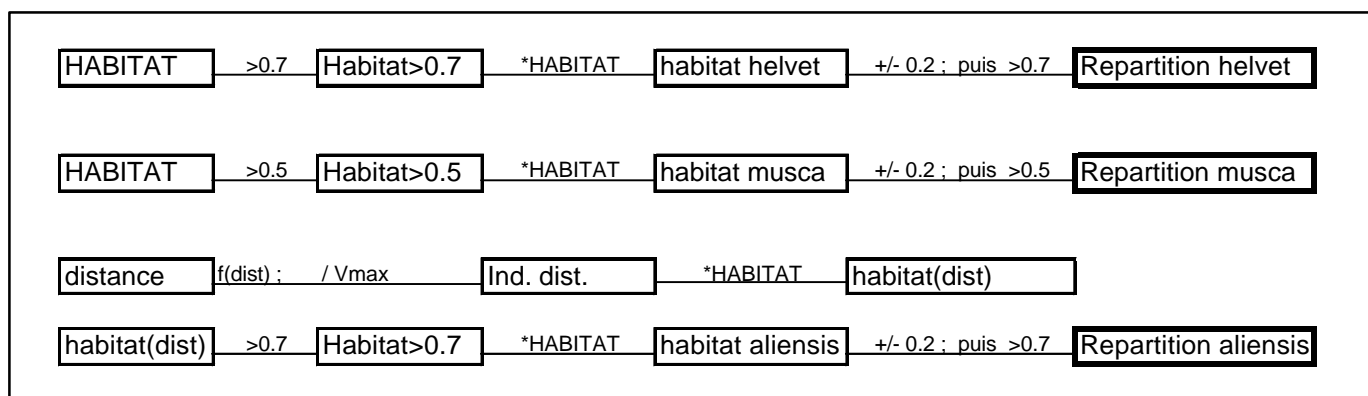
- Akçakaya, H.R. & Atwood, J.L. (1997). A habitat-based metapopulation model of the California Gnatcatcher. *Conservation Biology* **11**(2) : 422-434.
- Batem, Verhoest, E., Cambier, J.-L. (1991). *L'encyclopédie du Marsupilami de Franquin – La Grande Enigme*. Marsu productions S. A. M.
- Bonham-Carter, G. F. (1994). *Geographic Information Systems for Geoscientists : Modelling with GIS*. Love Printing Service Ltd, Ontario.
- Eastman, J. R. (1997). *Idrisi for Windows*. Clark University, Worcester.
- Fielding, A H. & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* **24**(1) : 38-49.
- Hausser, J. (1995). *Mammifères de Suisse*. Bâles, Birkhäuser.
- Hirzel, A. H., Hausser, J., Perrin N. (2000). *Biomapper 1.0 beta*. Laboratory for Conservation Biology, University of Lausanne.
- Hirzel, A. H., Hausser, J., Perrin N. (in prep). Ecological-Niche Factor Analysis : a new tool to compute habitat-suitability maps. Laboratory for Conservation Biology, Institute of Ecology, University of Lausanne.
- Maggini, R. (1999). Etude de la distribution de *Formica exsecta* Nyl. au Parc National Suisse à l'aide d'un SIG. Institute of Zoology and Animal Ecology. Lausanne, Lausanne.
- Manel, S., Dias, J. M., Buckton, S.T. & Ormerod, S. J. (1999). Alternative methods for predicting species distribution : an illustration with Himalayan river birds. *Journal of Applied Ecology* **36** : 734-747.
- , (1999). *S-Plus 2000, Professional Edition for windows, Release 1*. Mathsoft, Inc.
- Métral, F. & Helfer, V. (2000). *Marsupimapper*. Laboratory for Conservation Biology, Institute of Ecology, University of Lausanne.
- Sokal, R. R. & Rohlf, F. J. (1981). *Biometry : The principles and practice of statistics in biological research*. New York, W. H. Freeman & co.
- Zimmermann, F. (1998). Dispersion et survie des Lynx (*Lynx lynx*) subadultes d'une population réintroduite dans la chaîne du Jura. *Kora Bericht Nr. 4*, Muri (CH).

### Création de la carte de qualité d'habitat



Les encadrés représentent des cartes raster. La lettre "P" (probabilité) signifie qu'il s'agit d'un indice de qualité d'habitat, « .fq » d'une fréquence. Le chiffre en premier caractère indique le poids donné à la carte pour la création de la carte de qualité d'habitat.  $\mu$  : moyenne,  $\sigma$  : écart-type. La transformation « -lacs » correspond à une multiplication par une carte booléenne où lac=0. Le symbole « +/- 0.05 » indique l'introduction de stochasticité par addition de nombres aléatoires compris entre -0.05 et +0.05.

### Répartition des trois sous-espèces



Les encadrés représentent des cartes raster. Les espèces *M. musca-cerevisiae* et *M. helveticus* ont été traitées de façon similaires, une opération supplémentaire a été effectuée pour *M. aliensis*

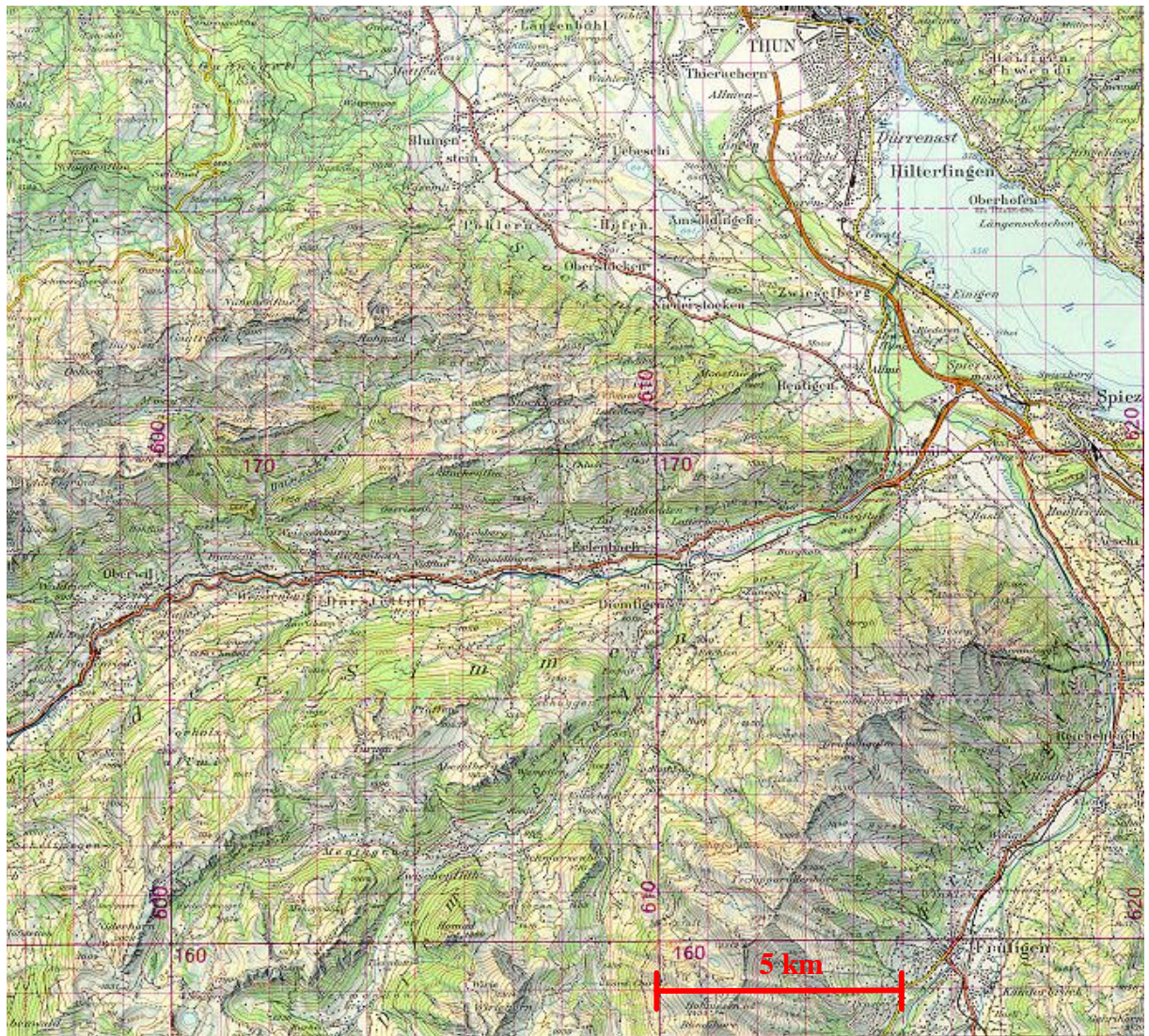
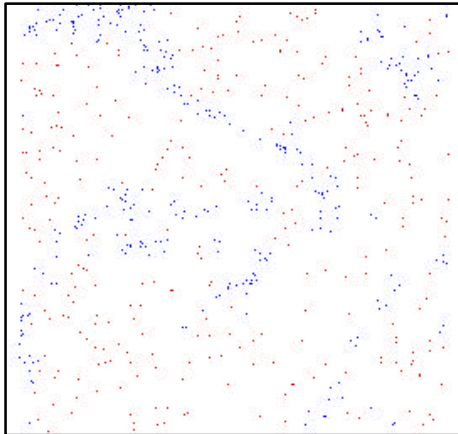
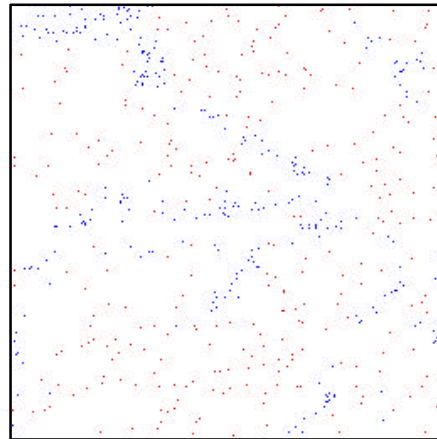


Figure a : Carte de la région choisie pour les analyses

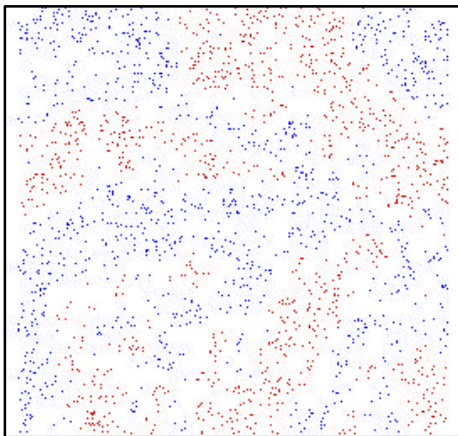




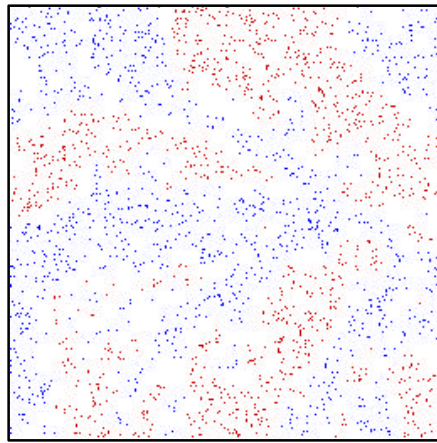
**Figure b:** Carte de calibration pour *helveticus*  
Nb. de présences : 305  
Nb. d'absences : 319



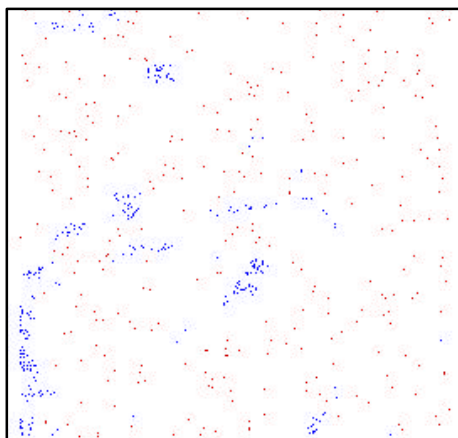
**Figure e:** Carte de validation pour *helveticus*  
Nb. de présences : 269  
Nb. d'absences : 268



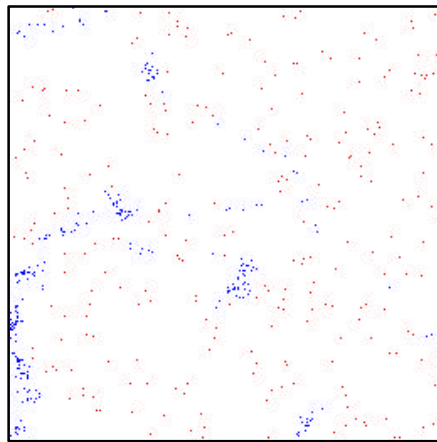
**Figure c:** Carte de calibration pour *musca-cerevisiae*  
Nb. de présences : 1100  
Nb. d'absences : 1063



**Figure f:** Carte de validation pour *musca-cerevisiae*  
Nb. de présences : 1095  
Nb. d'absences : 1128



**Figure d :** Carte de calibration pour *aliensis*  
Nb. de présences : 291  
Nb. d'absences : 301



**Figure g :** Carte de validation pour *aliensis*  
Nb. de présences : 249  
Nb. d'absences : 248

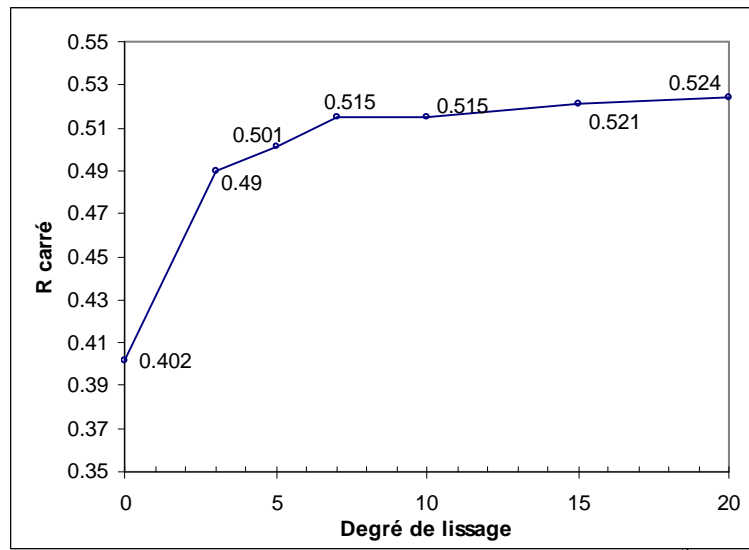


Figure h : Effet du lissage sur le coefficient de corrélation ( $R^2$ )

## Extrait des résultats obtenus pour l'analyse factorielle de la niche écologique sur *Biomapper*

### HELVETICUS

#### Valeurs propres :

	Valeur	Var. expliquée	Var. expl. cumulée
1	17.644	0.476	0.476
2	4.955	0.134	0.609
3	2.967	0.080	0.689
4	2.581	0.070	0.759
5	1.765	0.048	0.807
6	1.613	0.043	0.850
7	1.273	0.034	0.884
8	1.261	0.034	0.918
9	1.010	0.027	0.946
10	0.836	0.023	0.968
11	0.617	0.017	0.985
12	0.562	0.015	1.000

#### Scores

Variables	Facteurs								
	1	2	3	4	5	6	7	8	9
ALTIT.BOX.QT	-0.330	-0.127	0.019	-0.001	0.044	0.004	-0.005	0.007	-0.002
DFORETS.BOX.QT	-0.528	0.178	-0.012	-0.040	0.008	0.031	-0.016	-0.009	0.006
DLACS.BOX.QT	0.317	-0.092	-0.090	-0.018	0.013	0.042	-0.004	0.010	-0.010
DPAT.BOX.QT	0.321	0.059	-0.010	-0.022	0.069	-0.045	-0.006	0.001	0.011
DPRES.BOX.QT	-0.234	0.085	0.055	0.010	0.015	0.037	0.023	0.001	0.000
DRIV.BOX.QT	-0.377	0.037	0.018	0.012	0.005	-0.013	-0.044	0.004	-0.037
DROUTE1.BOX.QT	-0.137	0.007	0.029	0.039	-0.041	-0.015	-0.020	0.010	0.008
DVILLE.BOX.QT	0.043	-0.032	0.045	0.050	0.009	-0.011	0.018	-0.012	0.002
EXPO.BOX.QT	-0.095	0.044	0.011	-0.027	-0.007	-0.013	0.020	0.063	-0.008
FQBROU.BOX.QT	-0.262	-0.058	0.034	-0.117	-0.027	-0.019	0.004	-0.017	-0.002
FQROCH.BOX.QT	-0.320	0.053	-0.110	0.020	-0.018	-0.033	0.015	-0.007	0.001
PENT30.BOX.QT	-0.107	-0.035	-0.001	-0.004	-0.006	0.008	-0.032	0.018	0.039

**MUSCA-CEREVISIAE****Valeurs propres :**

	Valeur	Var. expliquée	Var. expl. cumulée
1	5.223	0.248	0.248
2	4.257	0.202	0.450
3	2.074	0.098	0.548
4	1.442	0.068	0.616
5	1.405	0.067	0.683
6	1.286	0.061	0.744
7	1.149	0.054	0.799
8	1.096	0.052	0.851
9	1.002	0.048	0.898
10	0.789	0.037	0.935
11	0.741	0.035	0.971
12	0.619	0.029	1.000

**Scores**

Variables	Facteurs						
	1	2	3	4	5	6	7
ALTIT.BOX.QT	-0.244	0.506	0.083	-0.112	0.059	-0.048	0.039
DFORETS.BOX.QT	-0.552	-0.275	0.006	-0.007	0.037	0.108	0.060
DLACS.BOX.QT	0.342	0.360	-0.117	0.000	0.108	0.058	0.029
DPAT.BOX.QT	0.251	-0.251	-0.071	0.026	0.127	-0.163	0.040
DPRES.BOX.QT	-0.281	-0.146	0.043	-0.122	0.006	0.017	0.015
DRIV.BOX.QT	-0.321	-0.144	0.034	0.006	-0.022	-0.079	0.089
DROUTE1.BOX.QT	-0.113	0.022	0.023	0.004	-0.116	-0.034	0.007
DVILLE.BOX.QT	0.185	0.107	0.055	-0.086	-0.093	-0.075	0.004
EXPO.BOX.QT	-0.068	-0.172	0.040	-0.093	0.097	-0.018	-0.152
FQBROU.BOX.QT	-0.275	0.112	0.299	0.145	0.050	-0.010	-0.027
FQROCH.BOX.QT	-0.378	0.207	-0.185	0.100	-0.033	-0.050	-0.068
PENT30.BOX.QT	-0.049	0.026	-0.002	0.030	0.064	0.019	0.063

**ALIENSIS**

**Valeurs propres :**

	Valeur	Var. expliquée	Var. expl. cumulée
1	19.575	0.392	0.392
2	11.140	0.223	0.615
3	5.996	0.120	0.735
4	3.017	0.060	0.796
5	2.932	0.059	0.854
6	1.590	0.032	0.886
7	1.386	0.028	0.914
8	1.193	0.024	0.938
9	1.097	0.022	0.960
10	0.818	0.016	0.976
11	0.721	0.014	0.991
12	0.473	0.009	1.000

**Scores**

Variables	Facteurs				
	1	2	3	4	5
ALTIT.BOX.QT	-0.235	0.335	0.032	-0.021	-0.053
DFORETS.BOX.QT	-0.421	-0.226	0.027	-0.029	-0.042
DLACS.BOX.QT	0.279	0.248	-0.108	-0.057	-0.015
DPAT.BOX.QT	0.243	-0.086	-0.013	-0.038	0.040
DPRES.BOX.QT	-0.242	-0.094	0.029	0.045	-0.009
DRIV.BOX.QT	-0.444	-0.097	0.056	0.018	-0.011
DROUTE1.BOX.QT	-0.195	0.116	-0.010	0.074	0.015
DVILLE.BOX.QT	0.453	0.155	-0.005	0.075	0.050
EXPO.BOX.QT	0.036	-0.114	0.042	-0.004	-0.042
FQBROU.BOX.QT	-0.235	0.085	0.216	-0.050	0.078
FQROCH.BOX.QT	-0.271	-0.115	-0.164	-0.023	0.070
PENT30.BOX.QT	-0.045	0.050	-0.029	-0.035	0.018