



Rapport Statistique

Aurore COMTE

Master 2 – Génétique et Biodiversité



I. Problématique

Mastomys natalensis et *Mastomys erythroleucus* sont deux espèces jumelles de rongeurs africains. Il a été montré au cours de deux expériences que ces animaux s'hybridaient en animalerie, mais pas en semi-liberté. On se demande alors quelles sont les barrières empêchant l'hybridation en semi-liberté, comment elles devraient varier pour permettre une hybridation, et si une introgression est possible sur le long terme.

Les résultats de l'expérience en semi-liberté nous apportent des éléments pour quantifier les paramètres des barrières pré-zygotiques. Il s'agit d'une expérience en enclos (enclos de 20m x 20m, voir annexe). Au début de l'expérience, 6 *M. erythroleucus* (mâles et femelles) ont été introduits dans un enclos et 6 *M. natalensis* (mâles et femelles) dans un autre. Les deux enclos étaient séparés par un système de portes qui ont été ouvertes après 3 mois, laissant circuler les rongeurs de l'un à l'autre. Cette expérience a duré 13 mois au total et a été réalisée en deux réplicats (2 x 2 enclos). L'étude de l'évolution de la population de l'enclos au cours du temps a été réalisée par capture-marquage-recapture avec 20 pièges installés pendant 4 jours dans chaque enclos au cours de sessions mensuelles (15 sessions au total). L'espèce de chaque nouvel individu capturé était reconnue grâce à un typage microsatellite effectué sur l'ADN extrait et amplifié à partir des phalanges récoltées sur chaque individu lors de son marquage.

L'analyse des données obtenues pendant les 15 sessions de capture-marquage-recapture vont pouvoir nous apporter des informations sur l'évolution de la répartition dans l'espace des deux espèces. Une espèce explore-t-elle plus l'espace que l'autre ? Les deux espèces s'évitent-elles ? Leurs territoires se chevauchent-ils ?

II. Test statistique préliminaire

A. Méthode

Nous allons chercher à déterminer si les rongeurs sont répartis au hasard dans l'enclos. Les pièges sont répartis de façon uniforme. Chaque piège peut donc être assimilé à un quadrat (une part de l'enclos découpé en parts égales) représentatif de la portion d'enclos dans laquelle il se situe.

La répartition régulière des individus dans l'enclos correspond à la présence d'un nombre identique d'individus par quadrat. A cause des erreurs d'échantillonnage, ces nombres ne seront cependant pas identiques et, si nous tracions le graphe représentant la fréquence des différents effectifs nous obtiendrions une courbe ressemblant à une loi binomiale positive. La moyenne d'une telle loi est toujours supérieure à la variance. Le rapport s^2/m (avec s^2 la variance et m la moyenne de la variable aléatoire « nombre d'individus par quadrat ») est donc inférieur à 1 pour une répartition régulière. Dans le cas d'une agrégation, nous aurions au contraire une loi binomiale négative et un indice s^2/m supérieur à 1. Si la répartition est au hasard, la variable « nombre d'individus par quadrat » suit une loi de Poisson avec $s^2/m = 1$. Le rapport s^2/m est un indice d'hétérogénéité ou d'agrégation spatiale (Downing, 1991). Notre hypothèse nulle (H_0) correspond à une répartition au hasard des individus dans l'enclos (loi de Poisson).

Par ailleurs, $s^2/m \times (n-1)$ correspond à une variable de test suivant approximativement une loi du khi deux avec $n-1$ degrés de liberté (Canard et Poinso, 2004). Le Khi-deux est ici utilisable car le nombre de pièges (/quadrat) est supérieur à 30 dans chaque enclos. Sous ces conditions, les 2,5% les plus élevés de la distribution correspondent à un écart à H_0 par excès d'hétérogénéité (binomiale positive) alors que les 2,5% les plus bas correspondent à un écart à H_0 par un excès d'homogénéité (binomiale négative). Entre ces deux bornes, on ne rejette pas H_0 (loi de poisson). Pour réaliser les tests, on effectue un changement de variable :

$Y = \sqrt{2\chi^2}$ qui suit une loi du chi-deux et est approximativement normale, de variance 1 (réduite) et de moyenne $\sqrt{2ddl - 1}$.

On centre cette nouvelle variable en retranchant sa moyenne pour obtenir une variable normale centrée réduite. Avec n le nombre de quadrats on obtient Z la valeur de cette loi :

$$Z = \sqrt{2\chi^2} - \sqrt{2n - 3}$$

Cette variable permet alors le test Z classique : si $|Z| > 1,96$, on rejette l'hypothèse H_0 « la répartition est aléatoire » au risque $\alpha = 0,05$.

Pour déterminer la raison du rejet éventuel, nous devons consulter l'indice $I = s^2/m$:

- Si $I < 1$, alors H_0 a été rejetée pour cause de répartition régulière.
- Si $I > 1$ alors H_0 a été rejetée pour cause de répartition agrégative.

Nous nous attendons à rejeter l'hypothèse H_0 pour cause de de répartition agrégative aux cours des 3 premières sessions lorsqu'on étudie séparément les données des *M. erythroleucus* et des *M. natalensis*. En effet, chaque espèce est d'un côté de l'enclos, il y a donc agrégation. Dans les sessions avec les portes ouvertes, nous nous attendons à rejeter moins souvent l'hypothèse nulle car les rongeurs ont accès à tout l'enclos.

B. Résultats

Dans le cas des deux espèces, nous présentons tout d'abord (figure ci-dessous) les valeurs respectives du chi-deux pour chaque session pour les deux réplicats et les deux espèces.

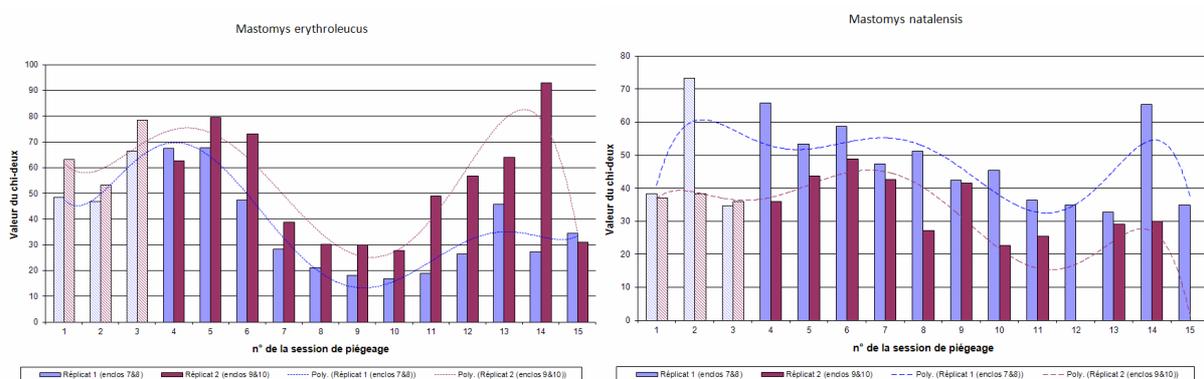


Figure 1 : graphiques de l'évolution de la valeur du chi-deux en fonction des sessions pour la répartition des *M. erythroleucus* et *M. natalensis*. Les trois histogrammes plus clairs correspondent aux sessions 1, 2 et 3 où les portes n'étaient pas encore ouvertes entre les enclos. Les courbes correspondent à une interpolation polynomiale visant à montrer l'évolution concomitante de tendances.

On observe que les valeurs du chi-deux obtenues à chaque réplicat suivent la même tendance. On peut donc supposer que cette tendance n'est pas due au hasard et que les rongeurs suivent bien une dynamique caractéristique au sein de l'enclos.

La figure 2 présente la valeur absolue de la variable centrée réduite Z pour les différentes sessions des deux réplicats et pour les deux espèces. Elle nous permet également de voir la fréquence du rejet ou non de l'hypothèse H_0 .

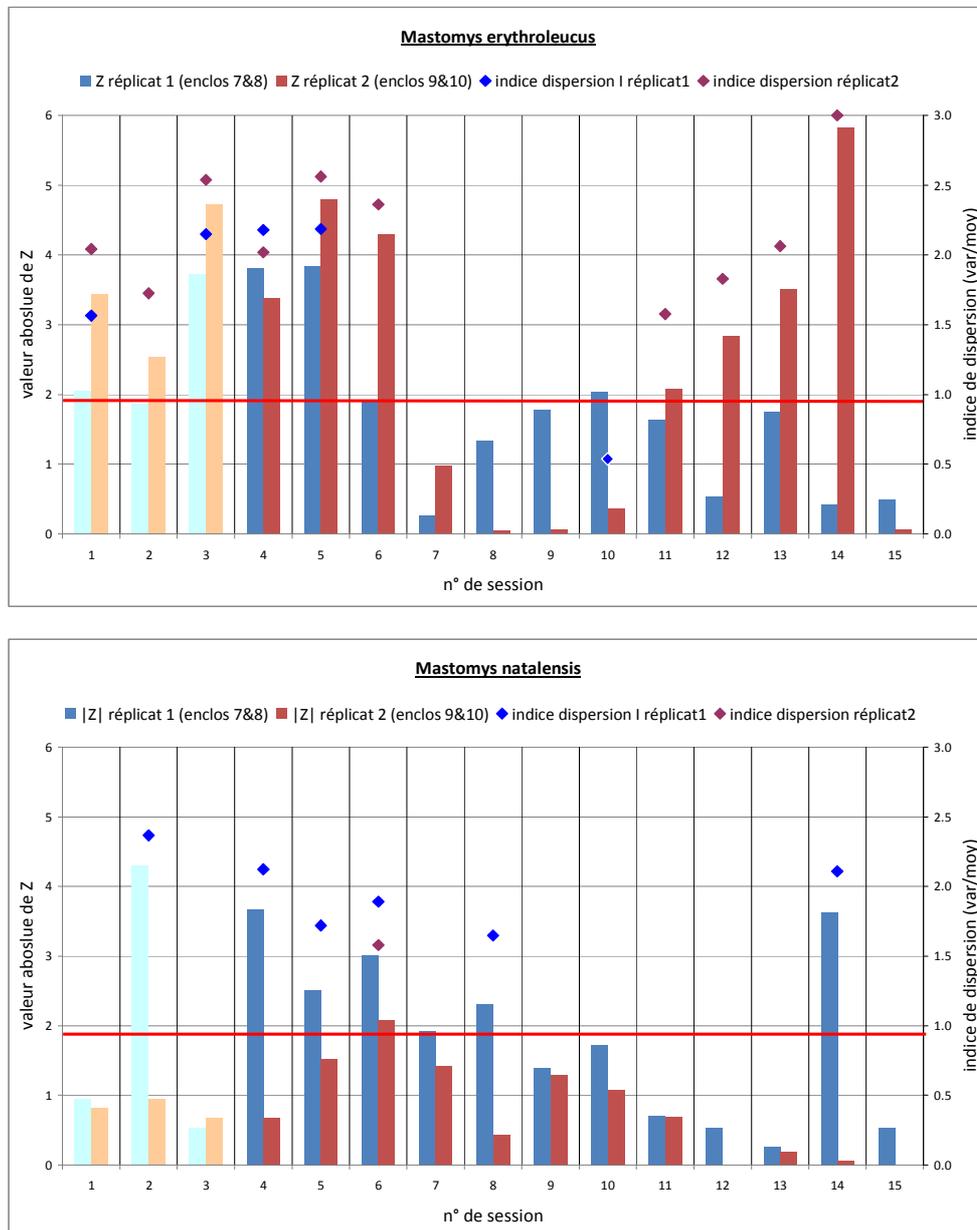


Figure 2 : graphique de l'évolution de la valeur absolue de Z en fonction des sessions pour la répartition des *M. erythroleucus* (au-dessus) et des *M. natalensis* (en dessous). Les trois histogrammes plus clairs correspondent aux sessions 1, 2 et 3 où les portes n'étaient pas encore ouvertes entre les enclos. La ligne rouge correspond à $|Z|=1.96$, soit la valeur à laquelle on peut rejeter l'hypothèse avec une erreur de 5%. Les points correspondent à la valeur de l'indice de dispersion en fonction des sessions, pour les sessions où Z est significatif.

Pour les *Mastomys erythroleucus*, on observe sur la figure 2 que la répartition des individus est agrégative dans l'enclos jusqu'à la session 6. Puis, elle devient aléatoire de la session 7 à la session 10 où elle est homogène (H_0 rejetée, $I < 1$), et redevient agrégée pour les sessions 11 à 14.

Pour les sessions 4 à 6, on peut penser que les deux espèces de *Mastomys* ne se précipitent pas vers les portes ouvertes et que la phase d'exploration n'est pas immédiate. Il y a donc une phase où, même les portes ouvertes, les *Mastomys* restent dans leur enclos respectif. La phase d'exploration de l'autre moitié de l'enclos se déroulerait aux sessions 7, 8 et 9 dans lesquels il n'y aurait pas d'organisation précise des *Mastomys* dans l'enclos (homogénéité). Puis, dans les dernières sessions, des agrégations de rongeurs se reforment, les rongeurs s'organisent.

Sur la figure 2, on observe que pour les *Mastomys natalensis*, le schéma est à peu près le même que celui des *Mastomys erythroleucus*. On observe par ailleurs que les 3 premières sessions nous suggèreraient une homogénéité des *M. natalensis* dans l'enclos, or ceci n'est pas possible car les deux enclos cloisonnés sont pris en compte dans les calculs. On peut penser que c'est un biais dû au faible nombre de rongeurs au début de l'expérience.

III. Visualisation des données

Le test statistique préliminaire ne permet pas de représenter plus en détail l'évolution des répartitions et les comportements concomitants de chaque espèce l'une par rapport à l'autre dans l'enclos. Nous avons donc cherché à représenter la dynamique spatiale du comportement de chaque espèce.

A. Méthode

Tout d'abord, nous avons attribué des coordonnées à tous les pièges afin de pouvoir utiliser les données dans R (figure 6 de l'annexe). Puis, grâce à un tableau croisé dynamique, le nombre de rongeurs présents à chaque piège en fonction de leur espèce a pu être compté. Un tableau pour chaque session a été obtenu (voir annexe).

Une technique couramment utilisée pour l'analyse d'une distribution spatiale d'individus est celle du « Kernel Smoothing » (Hengl et al., 2009). Cette technique pondère l'influence de plusieurs points les uns sur les autres en fonction de leur proximité (Matthiopoulos, 2001). La densité est estimée à chaque observation en utilisant une fonction appelée Kernel, qui peut-être par exemple une distribution gaussienne. Chaque fonction kernel est centrée sur une des observations. Une estimation de la densité totale peut être trouvée en sommant le poids des densités de chaque Kernel à chaque observation (Zieffler et al., 2001).

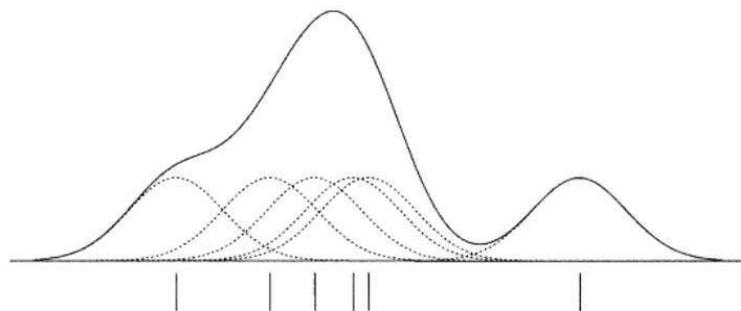


Figure 3 : Illustration de l'estimation de la densité (ligne pleine) pour $n=6$ observations (une observation correspond à un trait vertical sous la figure). Les lignes en pointillés correspondent aux fonctions kernel gaussiennes, à paramètre de lissage fixé, centrées sur chaque observation (extrait de Zieffler et al., 2001).

La méthode a été mise en œuvre en utilisant le package Spatstat de R (<http://www.spatstat.org/>, Baddeley 2005). Dans notre cas, nous avons un nuage de points et deux dimensions. Chaque point correspond à un piège et est pondéré de différentes façons selon le nombre d'individus pris dans ce piège. La création de Kernel à chaque point va nous permettre d'observer le gradient de densité dans l'enclos.

B. Résultats

La figure 4 nous montre l'évolution dans le temps de la distribution de *M. natalensis* et de *M. erythroleucis* pour les deux répliquats issue de l'application de la méthode du 'kernel smoothing' pour chaque session, répliquat et espèce de l'expérimentation.

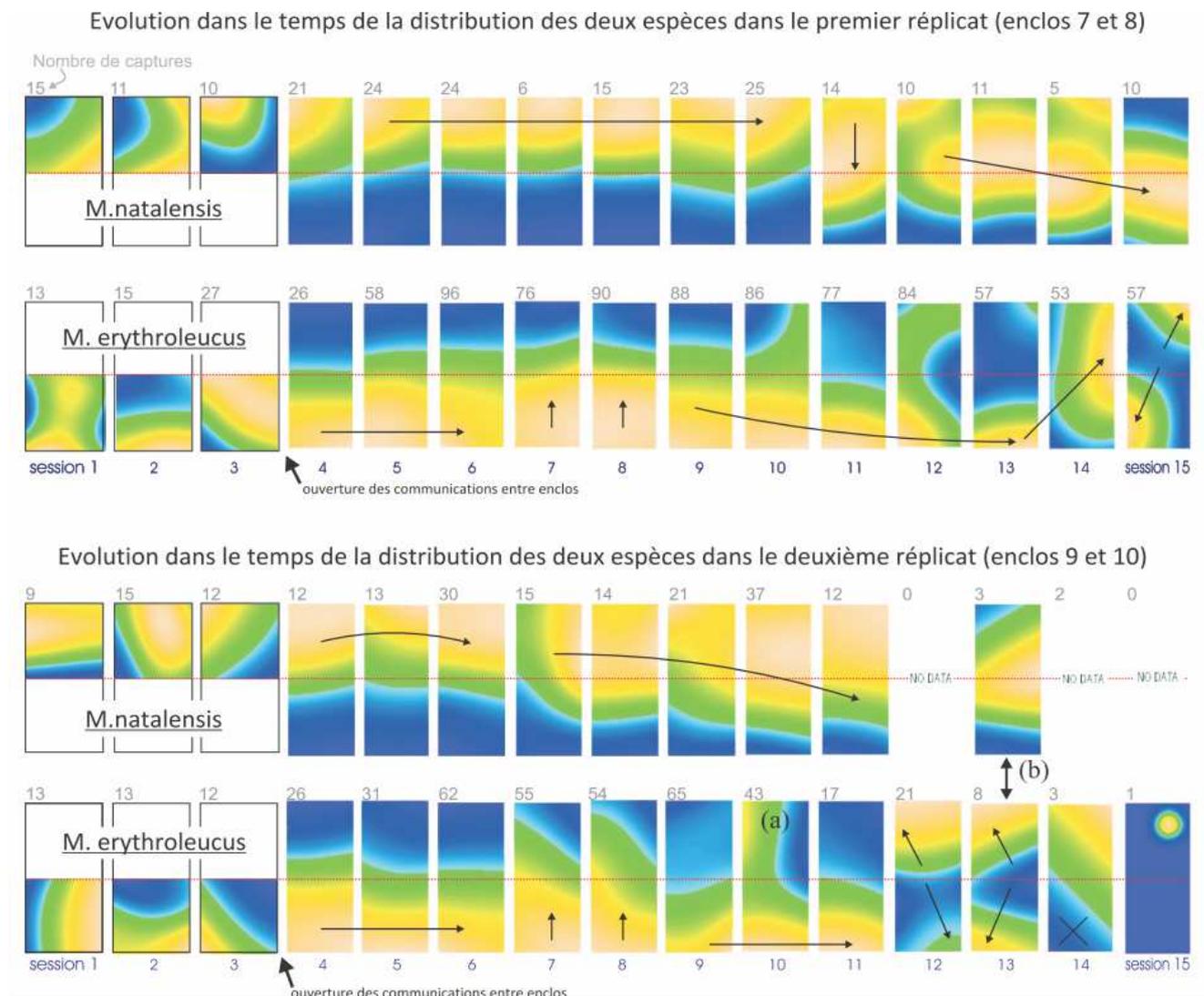


Figure 4: Visualisation du gradient de densité de la répartition des *M. natalensis* et *M. erythroleucis* grâce à la méthode du « Kernel Smoothing ». Les deux répliquats (enclos 7-8 et enclos 8-9) ont été traités de façon identique. Le jaune correspond aux zones où il y a une densité importante de rongeurs et le bleu là où il y en a une densité faible, voire nulle. Les flèches soulignent qualitativement les dynamiques des rongeurs. Le nombre de captures correspond au nombre d'individus trouvés dans les pièges à une session donnée, et donc au nombre d'individus utilisés pour réaliser ce « Kernel Smoothing ». Plus ce nombre de captures est élevé et plus la représentation est fiable (voir effectifs de capture en annexe).

La figure 4 nous permet d'observer que dans les 3 premières sessions (portes fermées), les individus semblent se répartir au hasard. Dans les sessions suivantes, comme nous l'avons vu au cours du test préliminaire, on observe une phase où les positions des espèces ne bougent pas. Puis, à partir de la neuvième session environ, on observe des mouvements dans les enclos pour les deux espèces. Lorsque les deux espèces sont en contact elles semblent se positionner l'une par rapport à l'autre. Leurs territoires semblent complémentaires dans l'enclos, et ils ne se superposent que très rarement sur des zones très petites. **On peut donc penser que les deux espèces s'évitent.** Les deux exemples de la figure 5 montrent bien ce phénomène ici.

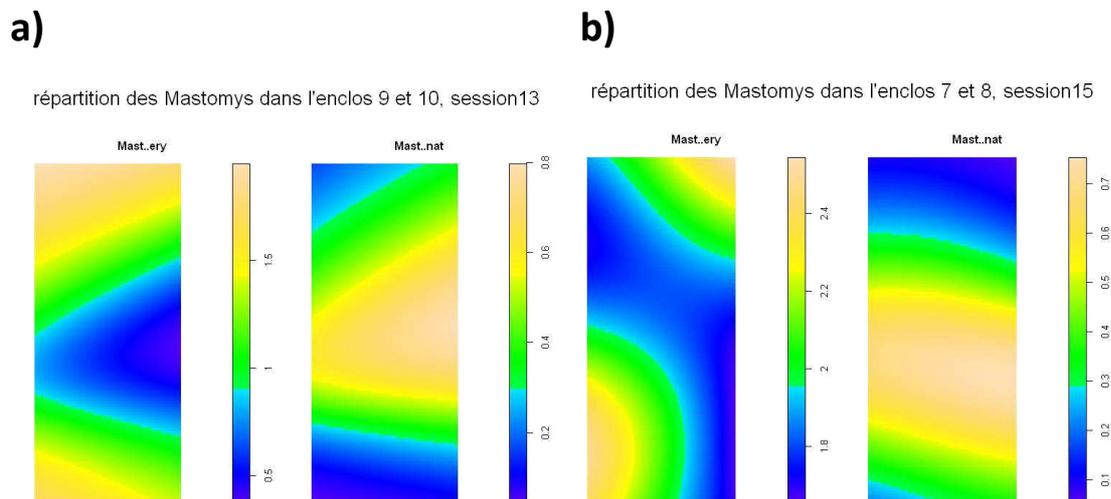


Figure 5: Gradient de densité correspondant à la répartition des *Mastomys natalensis* (gauche) et des *Mastomys erythroleucus* (droite) a) dans l'enclos 9-10 session 13. b) dans l'enclos 7-8 session 10.

Dans ces deux graphiques, on observe clairement deux patches de *M. erythroleucus*, séparés par un patch de *M. natalensis*, et les zones communes sont très réduites (figure 5 b), voire pratiquement inexistantes (figure 5 a). Cette observation pourrait-être l'une des explications de l'absence d'hybridation dans l'expérience en semi-liberté.

IV. Conclusion

Dans le modèle en cours de réalisation, nous avons prévu de modéliser la reconnaissance et l'attraction d'un individu d'une espèce par rapport à ses congénères du sexe opposé, et son indifférence par rapport aux individus d'une espèce différente. Or, d'après les résultats de ces tests, nous constatons que les barrières pré-zygotiques ne se limitent pas à cela. La reconnaissance entre deux individus de deux espèces différentes, quel que soit leur sexe, entraîne une répulsion si bien que les territoires des deux espèces se chevauchent le moins possible, limitant les rencontres et donc l'hybridation. Quelques situations (a) sur la figure 4 existent cependant, où des individus des deux espèces cohabitent, et constituent des contextes potentiellement favorables à une hybridation.

Bibliographie

Baddeley, A. (2008) *Analysing spatial point patterns in R (Version 3)*. The Commonwealth Scientific and Industrial Research Organisation.

Baddeley, A. (2010) *Analysing spatial point patterns in R (Version 4.1)*. The Commonwealth Scientific and Industrial Research Organisation.

Baddeley, A. and Turner, R. (2005) spatstat: An R Package for Analyzing Spatial Point Patterns. *Journal of Statistical Software*, 12 (6).

Broennimann, O., Fitzpatrick, C., M., Pearman, B., P., Petitpierre, B., Pellissier, L., Yoccoz, G., N., Thuiller, W., Fortin, M.-J., Randin, C., Zimmermann, E., N., Graham, H., C., and Guisan, A. (2012) Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecology and Biogeography*, 21 : 481–497.

Canard, A. and Poinot, D. (2004) Quelques méthodes statistiques typiques de l'étude des populations et des peuplements par la méthode des quadrats.

Diggle, J., P. (2003) *Statistical analysis of spatial point patterns*. Arnold.

Downing, JA (1991) Biological heterogeneity in aquatic ecosystems. Pages 160-180 in J. Kolasa and S.T.A. Pickett (editors). *Ecological heterogeneity*. Springer-Verlag, New York.

Froelich F., D. (2009) A note on the choice of the smoothing parameter in the kernel density estimate. *BRAC University Journal*, 1(1) : 59-68.

Hengl, T., Sierdsema, H., Radovic, A. and Dilo A. (2009) Spatial prediction of species' distributions from occurrence-only records: combining point pattern analysis, ENFA and regression-kriging. *Ecological Modelling* 220 : 3499–3511.

Klemelä, J. (2009) *Smoothing of Multivariate Data: Density Estimation and Visualization*. John Wiley & Son, New Jersey.

Matthiopoulos, J. (2001) *How to be a Quantitative Ecologist, the 'A to R' of Green Mathematics and Statistics*. John Wiley & Sons, United Kingdom.

Walter, J.-M., N. (2000-2001, mise à jour 2006) *La méthode de Ripley pour l'analyse des structures spatiales ponctuelles en écologie*.

Zieffler, S., A., Harring, R., J. and Long, D., J. (2001) *Comparing Groups: Randomization and Bootstrap Methods Using R*. John Wiley & Sons, New Jersey

Annexes

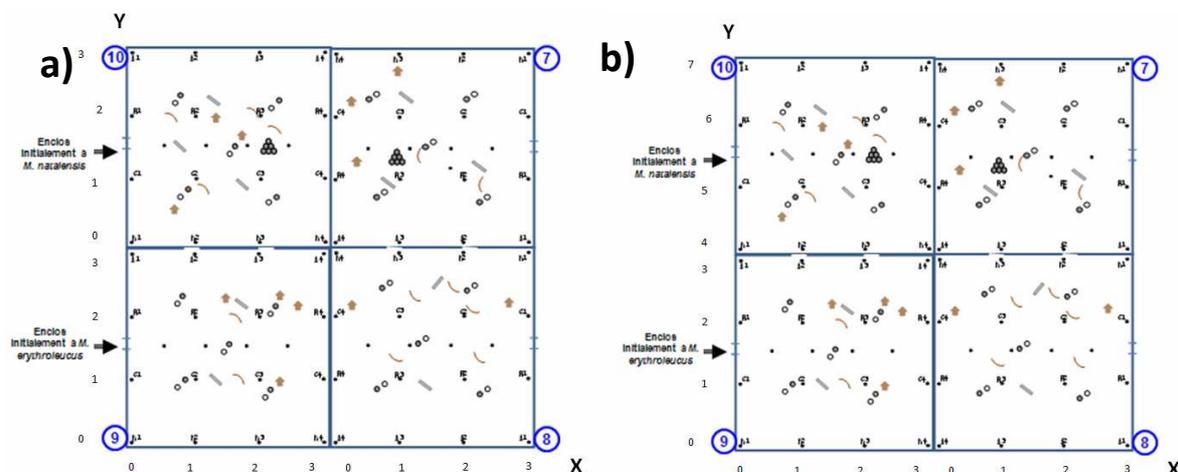


Figure 6 : mise en place d'axes et de coordonnées pour chaque piège. a) correspond aux coordonnées des pièges pour les quatre premières sessions où les portes sont fermées. On considère les deux réplicats indépendamment, et dans chaque réplicat, les deux enclos indépendamment. B) correspond aux onze dernières sessions. On considère les réplicats indépendamment, mais les portes sont ouvertes.

Tableau 1 : exemple de présentation des données avant transformation: session 1.

N°	Sexe	Poids(g) lâcher	Assignation cyt B	Session numéro 1: du 2 au 5 mars 2010				Observations	
				Poids/(g)	relevé 1	relevé2	relevé3		relevé4
1	F	41	Mast. nat	41	7B1	7A1	7A1	7B1	
2	F	38	Mast. nat	38		7B2	7B2	7B4	
3	F	44	Mast. nat	51	7D4	7B3	7C3		GEST
4	F	33	Mast. nat	29		7C1	7D1		
5	M	26	Mast. nat						
7	M	54	Mast. nat	55	7C2	7B4		7D1	
8	F	33	Mast. nat	30				10D2	
9	F	36	Mast. nat	42		10B3	10A3		
10	F	36	Mast. nat						
11	F	37	Mast. nat	36			10B2	10B2	
12	M	51	Mast. nat	43,5		10D3			
13	M	60	Mast. nat	59		10C3	10B3	10A2	
14	F	31	Mast. ery						
15	F	38	Mast. ery	67		8C2		8A1	GEST
17	F	42	Mast. ery	64	8B2				GEST
18	M	31	Mast. ery	41	8B1	8A3			
19	M	44	Mast. ery	46,5	8C2	8D2	8C2	8D4	
20	M	33	Mast. ery	43	8D4	8C4	8A1	8A3	
21	F	25	Mast. ery	41	9C3	9C4	9D4	9D4	GEST
22	F	36	Mast. ery	71	9B4	9B4	9B4	9A2	GEST, MIS BAS
23	F	40	Mast. ery	42,5	9C3				

24	M	41	Mast. ery	41	9C1				Trouvé MORT le 22/03/10
25	M	44	Mast. ery	50			9C3	9D2	
27	M	20	Mast. ery	26				9D4	
28	M		Mast. ery						

Tableau 2 : exemple de présentation des données après transformation: session 1.

zone	x	y	Piège	M. ery	M. nat	Total	zone	x	y	Piège	M. ery	M. nat	Total
10	0	3	10A1	0	0	0	8	3	0	8A1	2	0	2
10	1	3	10A2	0	1	1	8	2	0	8A2	0	0	0
10	2	3	10A3	0	1	1	8	1	0	8A3	2	0	2
10	3	3	10A4	0	0	0	8	0	0	8A4	0	0	0
10	0	2	10B1	0	0	0	8	3	1	8B1	1	0	1
10	1	2	10B2	0	2	2	8	2	1	8B2	1	0	1
10	2	2	10B3	0	2	2	8	1	1	8B3	0	0	0
10	3	2	10B4	0	0	0	8	0	1	8B4	0	0	0
10	0	1	10C1	0	0	0	8	3	2	8C1	0	0	0
10	1	1	10C2	0	0	0	8	2	2	8C2	3	0	3
10	2	1	10C3	0	1	1	8	1	2	8C3	0	0	0
10	3	1	10C4	0	0	0	8	0	2	8C4	1	0	1
10	0	0	10D1	0	0	0	8	3	3	8D1	0	0	0
10	1	0	10D2	0	1	1	8	2	3	8D2	1	0	1
10	2	0	10D3	0	1	1	8	1	3	8D3	0	0	0
10	3	0	10D4	0	0	0	8	0	3	8D4	2	0	2
7	3	0	7A1	0	2	2	9	0	3	9A1	0	0	0
7	2	0	7A2	0	0	0	9	1	3	9A2	1	0	1
7	1	0	7A3	0	0	0	9	2	3	9A3	0	0	0
7	0	0	7A4	0	0	0	9	3	3	9A4	0	0	0
7	3	1	7B1	0	2	2	9	0	2	9B1	0	0	0
7	2	1	7B2	0	2	2	9	1	2	9B2	0	0	0
7	1	1	7B3	0	1	1	9	2	2	9B3	0	0	0
7	0	1	7B4	0	2	2	9	3	2	9B4	3	0	3
7	3	2	7C1	0	1	1	9	0	1	9C1	1	0	1
7	2	2	7C2	0	1	1	9	1	1	9C2	0	0	0
7	1	2	7C3	0	1	1	9	2	1	9C3	3	0	3
7	0	2	7C4	0	0	0	9	3	1	9C4	1	0	1
7	3	3	7D1	0	2	2	9	0	0	9D1	0	0	0
7	2	3	7D2	0	0	0	9	1	0	9D2	1	0	1
7	1	3	7D3	0	0	0	9	2	0	9D3	0	0	0
7	0	3	7D4	0	1	1	9	3	0	9D4	3	0	3

Effectifs capturés lors des 15 sessions

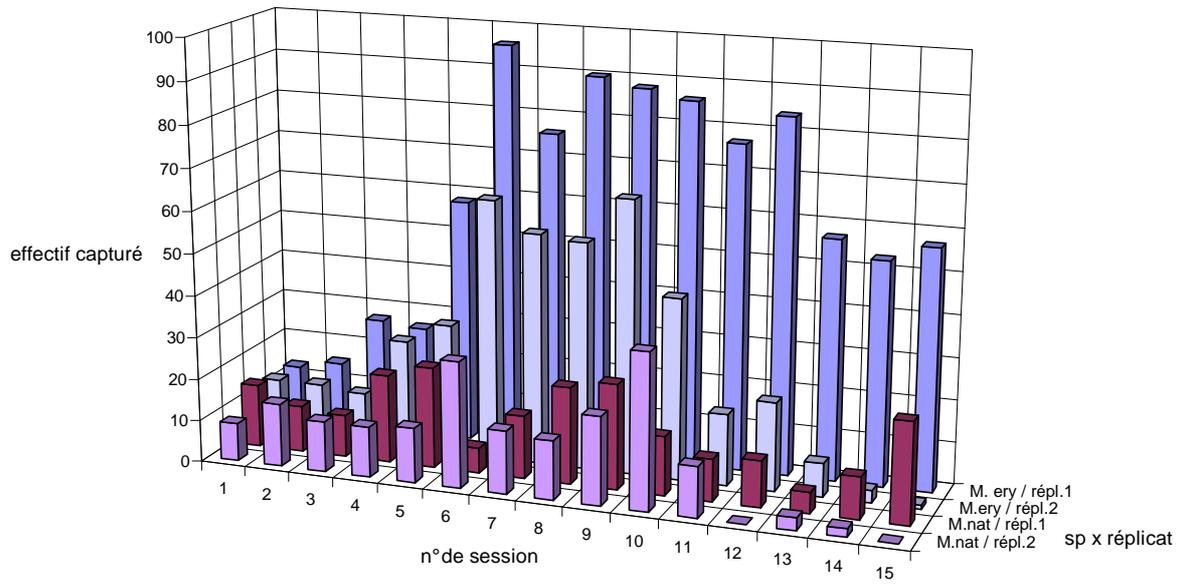


Figure 7 : nombre d'animaux capturés lors de chaque session.