

La répartition spatiale des *mustélidés* en France

Clément Calenge, Joël Chadoeuf, Christophe Giraud, Sylvie Huet, Romain Julliard, Pascal Monestiez, Jérémy Piffady, David Pinaud, Sandrine Ruetter

Publié: 26 mars 2015 • <https://doi.org/10.1371/journal.pone.0121689>

Abstrait

Nous avons estimé la distribution spatiale de 6 *Mustelidae* espèces en France à partir des données collectées par l'Agence Nationale de la Chasse et de la Faune Sauvage dans le cadre du programme «Journaux des petites espèces carnivores». Les 1500 agents nationaux de protection de la faune qui travaillent pour cette agence passent 80% de leur temps de travail à voyager dans la zone spatiale dans laquelle ils ont autorité. Au cours de leurs voyages, ils détectent occasionnellement des animaux carnivores morts ou vivants de petite et moyenne taille. Entre 2002 et 2005, chaque voiture exploitée par cette agence était équipée d'un journal de bord dans lequel les agents enregistraient des informations sur les animaux détectés (espèce, emplacement, morts ou vivants, date). Ainsi, plus de 30000 animaux morts ou vivants ont été détectés au cours de la période d'étude. Parce qu'un grand nombre d'animaux détectés dans une région aurait pu être le résultat d'une pression d'échantillonnage élevée là-bas, nous avons modélisé le nombre d'animaux détectés en fonction de l'effort d'échantillonnage pour permettre une estimation sans biais de la densité des espèces. Pour les animaux morts - majoritairement tués sur les routes - nous avons supposé que l'effort dans une région donnée était proportionnel à la distance parcourue par les officiers. Pour les animaux vivants, nous n'avions aucun moyen de mesurer l'effort d'échantillonnage. Nous avons démontré qu'il était possible d'utiliser l'ensemble des données (animaux morts et vivants) pour estimer ce qui suit: (i) la densité relative - c'est-à-dire la densité multipliée par une constante inconnue - de chaque espèce d'intérêt à travers les différents Français les régions agricoles, (ii) l'effort d'échantillonnage des animaux vivants pour chaque région, et (iii) la probabilité relative de détection pour diverses espèces d'intérêt. Pour les animaux morts - majoritairement tués sur les routes - nous avons supposé que l'effort dans une région donnée était proportionnel à la distance parcourue par les officiers. Pour les animaux vivants, nous n'avions aucun moyen de mesurer l'effort d'échantillonnage. Nous avons démontré qu'il était possible d'utiliser l'ensemble des données (animaux morts et vivants) pour estimer ce qui suit: (i) la densité relative - c'est-à-dire la densité multipliée par une constante inconnue - de chaque espèce d'intérêt à travers les différents Français les régions agricoles, (ii) l'effort d'échantillonnage des animaux vivants pour chaque région, et (iii) la probabilité relative de détection pour diverses espèces d'intérêt.

Référence: Calenge C, Chadoeuf J, Giraud C, Huet S, Julliard R, Monestiez P, et al. (2015) La distribution spatiale des *mustélidés* en France. PLoS ONE 10 (3): e0121689. <https://doi.org/10.1371/journal.pone.0121689>

Rédacteur académique: Cédric Sueur, Institut Pluridisciplinaire Hubert Curien, FRANCE

Reçu: 10 juillet 2014; **Accepté:** 18 février 2015; **Publié:** 26 mars 2015

Droits d'auteur: © 2015 Calenge et al. Il s'agit d'un article en libre accès distribué sous les termes de la [licence d'attribution Creative Commons](#), qui permet une utilisation, une distribution et une reproduction sans restriction sur tout support, à condition que l'auteur original et la source soient crédités.

Disponibilité des données: toutes les données pertinentes se trouvent dans le papier et ses fichiers d'informations complémentaires.

Financement: Ce document a été partiellement financé par: Le programme MASTODONS du Centre national de la recherche scientifique (<http://www.cnrs.fr/mi/spjp.php?article53&lang=fr>), sous le titre «Statistiques Crowdsourcing biodiversité» (auteur financé RR). Le programme CiSStats financé par le programme «réseaux incitatifs» de l'Institut national de la recherche agronomique (auteur financé PM, <http://ciam.inra.fr/cisstats/>). Le programme «Chaire de modélisation mathématique et biodiversité» de Véolia, École Polytechnique et Musée national d'histoire naturelle (auteurs financés CG et RJ; <http://www.cmap.polytechnique.fr/chaire-mmb/>). Les bailleurs de fonds n'ont joué aucun rôle dans la conception de l'étude, la collecte et l'analyse des données, la décision de publier ou la préparation du manuscrit.

Intérêts concurrents: les auteurs ont déclaré qu'il n'y avait pas d'intérêts concurrents.

introduction

Toutes les espèces de *Mustelidae* endémiques de France sont inscrites comme espèces strictement protégées à l'Annexe II (*Mustela lutreola*, *Lutra lutra*) ou en tant qu'espèce protégée à l'Annexe III de la Convention de Berne sur la conservation de la vie sauvage et du milieu naturel de l'Europe. Cependant, certaines de ces espèces sont toujours considérées comme nuisibles dans certaines régions des pays signataires, soit parce qu'elles peuvent être porteuses de maladies dangereuses pour l'homme et le bétail, soit parce qu'elles peuvent causer des dommages importants aux activités humaines. Pour ces raisons, la convention de Berne n'interdit pas la chasse et les autres formes d'exploitation de ces espèces protégées, tant que leurs populations sont à «un niveau qui correspond notamment aux exigences écologiques, scientifiques et culturelles» (article 2 de la convention), afin de «maintenir les populations hors de danger» (article 7).

La question de savoir si la densité d'une espèce donnée atteint le «niveau de population requis» dans un endroit donné est généralement déterminée par le gestionnaire de la faune. Il connaît généralement assez bien la situation de conservation de l'espèce dans quelques endroits, et il utilise ces endroits comme des «références» auxquelles d'autres endroits sont comparés pour évaluer la situation de l'espèce dans l'ensemble de la zone d'intérêt (tout en considérant d'autres mesures de la situation de l'espèce également, par exemple les dommages causés au bétail, le nombre d'animaux piégés). Étant donné que les estimations de densité sont couramment utilisées pour une comparaison entre différents endroits, il n'est généralement pas nécessaire de travailler avec des mesures de densité «absolues». Dans de nombreux cas, travailler avec un indice proportionnel à la densité réelle de l'espèce - que nous appelons «densité relative» - est suffisant pour une gestion rationnelle de la faune [1 , 2].

De plus, la densité relative est beaucoup plus facile à estimer que la densité absolue. D'une part, la plupart des méthodes statistiques développées pour l'estimation de la densité absolue reposent sur des conceptions encombrantes - par exemple, les méthodes de marquage-recapture, les approches d'échantillonnage à distance ou les modèles d'élimination [3 , 4]. De telles approches sont beaucoup trop coûteuses pour être considérées pour l'estimation de la densité de toutes les espèces de *Mustelidae* dans chaque région à l'échelle d'un pays entier. D'autre part, la densité relative peut être estimée à partir de conceptions beaucoup plus simples: un ensemble de sites est échantillonné au hasard dans la zone d'intérêt et des dénombrements d'organismes sont organisés pour ces sites. Sur un site donné, le dénombrement résultant peut être utilisé comme un proxy de la densité réelle (voir [5] à titre d'exemple). En effet, si l'hypothèse d'une détectabilité constante se vérifie dans l'espace et dans le temps, le nombre moyen d'animaux comptés par site échantillonné est proportionnel à la densité réelle de l'espèce dans la zone [1]. Ainsi, les estimations de la densité relative ont été couramment utilisées pour cartographier la distribution spatiale de plusieurs espèces [6].

Cependant, de tels plans d'échantillonnage développés pour l'estimation de la densité relative sont difficiles à mettre en œuvre pour les espèces de *Mustelidae* en France: ces espèces sont timides, et la plupart d'entre elles sont nocturnes et donc difficiles à détecter. L'hermine *Mustela ermina* et la belette *Mustela nivalis* sont en partie diurnes, mais ces deux espèces sont difficilement détectables car elles sont beaucoup plus petites que les autres *Mustelidae*. Enfin, comme pour de nombreuses espèces carnivores, leurs densités sont faibles par rapport aux espèces généralement suivies avec de tels modèles (par exemple, les oiseaux communs). Le suivi des espèces carnivores est donc une tâche complexe.

Pour cette raison, la plupart des auteurs de la littérature ont recommandé de surveiller les espèces carnivores à l'aide de ce que l'on appelle des «signes de terrain», tels que le comptage des traces [7] ou le comptage fécal [8 , 9]. Cependant, l'utilisation de panneaux sur le terrain est généralement associée à de graves problèmes d'identification des espèces. Par exemple, il est très difficile de déterminer avec certitude si une empreinte appartient à une hermine ou à une belette [10]; dans les études reposant sur la numération fécale, même des experts hautement qualifiés peuvent mal identifier une grande proportion de matières fécales [11]. De plus, l'utilisation d'un type donné de signe de champ peut être judicieuse pour une espèce, mais pas pour une autre (voir [10]). Enfin, les scores d'index basés sur les signes de terrain sont susceptibles d'être plus fiables lorsqu'ils sont utilisés pour estimer les tendances de la population au fil du temps dans un endroit que lorsqu'ils sont utilisés pour comparer entre les zones [12]. Pourtant, les panneaux de terrain sont généralement préférés aux autres approches reposant sur des dénombrements directs ou indirects d'animaux, comme les dénombrements par projecteurs [13] ou les statistiques de piégeage [14 , 15]. De tels dénombrements sont considérés comme impraticables car ils sont trop coûteux et pas assez sensibles pour une surveillance à grande échelle [16 , 17].

Cependant, nous montrons dans cet article que les dénombrements directs et indirects peuvent également être collectés à faible coût pour obtenir des estimations significatives de la densité relative, évitant ainsi - dans une certaine mesure (voir [discussion](#)) - la plupart des problèmes soulevés par l'utilisation de panneaux de terrain. Ainsi, en 2002–2005, l'Office national de la chasse et de la faune sauvage (Office national de la chasse et de la faune sauvage, ci-après ONCFS) a mis en place un plan de collecte de données d'occurrence sur les espèces de *Mustelidae*. Le programme des «journaux de bord des petites espèces carnivores» (ci-après le programme SCSL) reposait sur l'importante main-d'œuvre de cette organisation. En effet, les 1500 agents nationaux de protection de la faune travaillant pour cette organisation passent beaucoup de temps sur le terrain et détectent donc occasionnellement des *Mustelidae* vivants ou morts. Au cours de la période d'étude, chaque voiture officielle était équipée d'un journal de bord dans lequel les informations concernant chaque animal détecté par les agents de protection de la faune ont été enregistrées (c.-à-d. L'espèce, la municipalité, la date de détection et si l'animal était mort ou vivant).

Les données collectées n'ont pas pu être directement utilisées pour estimer les indices de densité relative des espèces d'intérêt dans les différentes régions françaises. En effet, le nombre d'animaux détectés dans une région donnée dépendait en partie du temps passé par les agents dans cette région. Un grand nombre d'animaux détectés dans une région donnée pourrait avoir été le résultat d'une abondance élevée de l'espèce ainsi que le résultat d'un effort d'observation intense (également appelé effort d'échantillonnage dans cet article). Un modèle du processus d'observation était nécessaire pour tenir compte de cette source de variation dans l'estimation d'abondance [18]. De plus, le processus d'observation n'était pas le même pour les animaux morts et vivants. Des animaux vivants ont souvent été détectés lors des opérations de jour des agents sur le terrain, tandis que des animaux morts ont souvent été détectés le long de la route (en raison de la mort sur la route) lors des déplacements des agents. Enfin, pour un statut donné (mort ou vivant), la détection variait probablement d'une espèce à l'autre en raison de leur variation de taille.

Dans cet article, nous avons utilisé les données collectées dans le cadre du programme SCSL pour estimer la densité animale moyenne dans des unités spatiales caractérisées par des pratiques et activités agricoles homogènes (petites régions agricoles) en France ([Fig. 1B](#)). Sur la base d'un modèle de l'effort d'échantillonnage des animaux morts, il a été possible d'estimer la densité relative de chaque espèce dans chaque région d'intérêt à l'aide d'un modèle log-linéaire. Nous avons également pris en compte la régularité spatiale attendue de la densité des espèces, ainsi que l'effet de l'environnement sur ces densités, en utilisant une régularisation de type crête. Nous fournissons un package R nommé `scsl` (voir [Package S1](#)). Ce package contient les données, le code C et les fonctions R utilisées pour les calculs effectués dans ce papier. Nous fournissons également un document rédigé à l'aide du système Knitr [19], décrivant l'ajustement et l'évaluation du modèle avec ces fonctions, ainsi que des analyses complémentaires effectuées pour démontrer la robustesse de nos résultats.

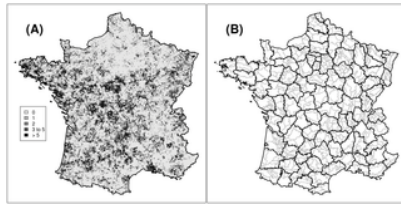


Fig 1. Données utilisées pour l'estimation de la densité relative des petites espèces carnivores en France.

(A) nombre d'animaux détectés dans chaque commune de France de 2002 à 2005 (l'encart contient la légende). (B) Carte des petites régions agricoles (SAR, en gris) et du département (en noir)

<https://doi.org/10.1371/journal.pone.0121689.g001>

Méthodes

Collecte de données

Les agents de protection de la faune, chargés à la fois de la protection de la faune et de la police de la chasse, sont des agents officiellement désignés mandatés par le ministre français de l'Écologie. La plupart d'entre eux sont affectés à un département (une unité administrative française, voir Fig.1B). Ces agents passent environ 80% de leur temps de travail sur le terrain. Ils mènent deux grands types d'activités: (i) des «activités statiques» au cours desquelles les agents passent beaucoup de temps hors de leur voiture dans une municipalité donnée (par exemple, surveillance de la faune, collecte de données pour la recherche, repérage pour arrêter les braconniers) et (ii) «activités de déplacement» au cours desquelles les agents passent beaucoup de temps à conduire leur voiture (par exemple, missions de patrouille de police). En général, une équipe donnée d'officiers (avec généralement deux officiers par équipe) conduit toujours la même voiture. De 2001 à 2005, le programme SCSL a profité de ce temps passé par les agents dans des activités statiques et mobiles.

Notre période d'étude couvre toutes les années de 2002 à 2005. Nous n'avons pas pris en compte la première année de suivi (2001) car le programme SCSL n'était pas encore bien implanté à ce moment. Nous avons supposé que les densités d'espèces étaient constantes pendant la période d'étude (une hypothèse raisonnable étant donné la courte période, mais voir la [discussion](#)). Au cours de notre période d'étude, chaque voiture de l'ONCFS était équipée d'un journal de bord collecté chaque année. Les 1500 agents de protection de la faune de cette organisation y ont enregistré chaque animal détecté appartenant aux espèces carnivores suivantes: (i) l'hermine *Mustela erminea*, (ii) la belette *Mustela nivalis*, (iii) le putois *Mustela putorius*, (iv) le blaireau *Meles meles*, (v) la martre des pins *Martes martes*, (vi) la martre *Martes foina*, (vii) la genette commune *Genetta genetta* et (viii) le chat sauvage *Felis silvestris*. A noter que les deux dernières espèces ne sont pas des *Mustelidae* et ne présentaient pas d'intérêt direct dans notre étude. Pour chaque animal détecté, les agents ont noté dans le journal de bord l'espèce, le statut de l'animal (mort ou vivant), la date de la détection et la commune où la détection a eu lieu. Lorsqu'un animal mort a été détecté plusieurs fois, une seule observation a été signalée. Ainsi, au cours de notre période d'étude, 31 811 animaux appartenant à ces 8 espèces ont été enregistrés (Fig.1A).

Compte tenu du processus de détection, deux grands groupes d'espèces peuvent être distingués. D'une part, en raison de leur petite taille, les hermines et les belettes étaient rarement vues mortes, car leurs corps se dégradent rapidement dans l'environnement ou disparaissent en raison du balayage. De plus, du fait qu'elles sont en partie diurnes, elles ont pu être détectées vivantes pendant la journée (plus de 75% des hermines et belettes ont en effet été détectées vivantes). Selon un sondage informel mené auprès d'agents de protection de la faune, des hermines et des belettes vivantes ont été détectées principalement lors d'activités statiques, et très rarement à partir d'une voiture en mouvement. En revanche, les autres espèces (blaireaux, martres, etc.) sont plus grosses et leurs corps mettent plus de temps à se dégrader. De plus, toutes ces espèces étant nocturnes, elles n'ont pu être détectées vivantes que lors des missions nocturnes des agents. Réellement, ces espèces ont été le plus souvent détectées par la route lors d'activités de déplacement (74% de ces animaux ont en effet été détectés morts). Nous avons supposé qu'il n'y avait pas d'erreurs d'identification (voir [discussion](#) sur cette question).

Pour 51% des années-voitures c , la distance k_c parcourue par la voiture au cours de l'année a également été reportée dans le journal de bord en fin d'année. Nous avons supposé que la probabilité de déclarer la distance parcourue par une voiture donnée ne dépendait pas de k_c (par exemple, les voitures très utilisées n'étaient pas plus susceptibles de déclarer leur distance parcourue que les voitures moins d'occasion). Sous cette hypothèse, et étant donné que toutes les voitures ont limité leurs mouvements à l'intérieur des limites du département dans lequel l'équipe d'officiers travaillait, il a été possible d'estimer grossièrement le nombre total de kilomètres T_d parcourus par toutes les voitures pendant la période d'étude dans un département d :

$$T_d = B_d \times \frac{1}{R_d} \sum_{c=1}^{R_d} k_c \quad (1)$$

où B_d est le nombre total de journaux de bord disponibles dans le département d pendant la période d'étude et R_d est le nombre de journaux de bord disponibles dans ce département pour lesquels la distance annuelle parcourue a également été enregistrée. Notons que la mesure T_d tient implicitement compte de la variation interannuelle du nombre de kilomètres parcourus par une voiture donnée, ces nombres de kilomètres étant additionnés au fil des années.

Cadre de modélisation de base

L'unité spatiale d'intérêt dans notre étude était la «petite région agricole». Les SAR sont définies par l'intersection entre la «grande région agricole» et le département. Chaque grande région agricole correspond à un groupe de communes voisines, appartenant éventuellement à des départements différents, caractérisés par une homogénéité des pratiques et activités agricoles (les 429 grandes régions agricoles sont définies par le ministère de l'Agriculture; pour plus de détails voir <http://www.agreste.agriculture.gouv.fr/définitions/zonages/>). L'intersection entre les 90 départements et les 429 grandes régions agricoles définit 703 petites régions agricoles (SAR, Fig.1B). Chaque département contient 7,8 SAR en moyenne (ET = 3,3). Les DAS sont

suffisamment petits et homogènes pour être utilisés à des fins de gestion de la faune, mais suffisamment grands pour garantir que la plupart d'entre eux (95%) contenaient au moins un animal détecté. Notre objectif était donc d'utiliser les données collectées dans le cadre du programme SCSL pour estimer la densité des 6 espèces de *Mustelidae* d'intérêt dans les 703 SAR de France.

Soit N_{ijk} le nombre total d'animaux de l'espèce i ($i = 1 \dots I$) dans le SAR j ($j = 1 \dots J$) de statut k ($k = 1$ correspondant aux animaux morts et $k = 2$ correspondant aux animaux vivants) détecté pendant toute la période d'étude (c'est-à-dire la somme, au cours des années de la période d'étude, des dénombrements annuels pour cette combinaison particulière i, j, k). Tout d'abord, nous avons supposé que N_{ijk} pouvait être décrit à l'aide d'une distribution de Poisson:

$$N_{ijk} \sim \mathcal{P}(\lambda_{ijk})$$

où λ_{ijk} est l'espérance de cette distribution. De plus, nous avons supposé le modèle suivant pour cette attente:

$$\lambda_{ijk} = S_j A_{ij} E_{jk} P_{ik} \quad (2)$$

où S_j est la surface connue du $j^{\text{ème}}$ SAR, A_{ij} est la densité réelle de l'espèce i dans le SAR j , E_{jk} est l'effort d'échantillonnage dans le SAR j pour les animaux du statut k , et P_{ik} est la probabilité de détection de l'espèce i dans le statut k . Notez que nous avons utilisé les termes «effort d'observation» et «effort d'échantillonnage» de manière interchangeable dans cet article. Ainsi, l'hypothèse principale à la base de ce modèle est que, pour un statut donné (animaux morts ou vivants), le processus d'observation peut être décomposé en deux composantes statistiquement indépendantes: un effort d'observation dépendant du site uniquement, et une probabilité de détection dépendant uniquement de les espèces.

Les termes «effort d'échantillonnage» et «probabilité de détection» ne sont ici que des conventions; le fait est que, pour un statut donné, aucune composante du processus d'observation ne variait à la fois selon les sites et selon les espèces. Par exemple, la détectabilité d'un animal vivant était probablement systématiquement plus faible dans un SAR avec une couverture dense de couverture herbacée que dans un SAR caractérisé par une forte couverture de sol nu. Cependant, le composant E_{jk} expliquerait ces différences de détectabilité entre les sites. D'autre part, cette hypothèse serait violée si certaines espèces étaient plus facilement détectables que d'autres dans certains sites, mais moins détectables dans d'autres sites. Cela pourrait se produire, par exemple, si l'attention portée à certaines espèces était plus élevée dans certaines zones que dans d'autres. Cependant, compte tenu de la grande homogénéité du réseau en termes de tâches et de formation, et de la simplicité du protocole (chaque observation aurait dû être enregistrée), nous avons jugé cela improbable.

Le modèle (2) appartient à la famille des modèles linéaires généralisés à interactions multiplicatives [20]. Comme nous avons formulé le problème ici, le modèle (2) est dégénéré en ce sens que tous les paramètres inconnus A_{ij} , E_{jk} et P_{ik} ne sont pas identifiables de manière unique. Une infinité de combinaisons de paramètres pourrait donner la même valeur de λ_{ijk} . Cependant, certains paramètres liés à l'effort d'échantillonnage sont effectivement connus. En effet, comme les animaux morts étaient essentiellement des tués sur la route détectés lors d'activités de déplacement, l'effort d'échantillonnage des animaux morts dans un SAR donné pourrait être considéré comme proportionnel au nombre total de kilomètres parcourus par les agents dans ce SAR. Comme nous le verrons plus loin, il est possible d'estimer ce nombre de kilomètres à partir des informations disponibles dans les journaux de bord. Par contre, nous n'avons aucune information sur l'effort d'échantillonnage des animaux vivants, qui était proportionnel au temps passé par les agents aux activités statiques - une quantité généralement inconnue.

Nous avons supposé par la suite que le nombre de kilomètres V_j parcourus par les officiers dans le $j^{\text{ème}}$ SAR était connu et proportionnel à l'effort d'échantillonnage réel E_{j1} pour les animaux morts, soit,

$$E_{j1} = \beta V_j$$

avec β une constante inconnue. Nous avons reparamétrisé le modèle:

$$s_j = \log S_j \quad (3)$$

$$a_{ij} = \log(\beta) + \log(P_{i1}) + \log(A_{ij}) \quad (4)$$

$$e_{jk} = \log\left(\frac{E_{jk}}{\beta}\right) + \log\left(\frac{P_{1k}}{P_{11}}\right) \quad (5)$$

$$p_{ik} = \log\left(\frac{P_{ik}}{P_{1k}}\right) + \log\left(\frac{P_{11}}{P_{11}}\right) \quad (6)$$

où P_{1k} est la probabilité de détection de l'espèce 1 (ici, la belette) dans le statut k , P_{i1} est la probabilité de détection des animaux morts de l'espèce i , et P_{11} est la probabilité de détection des belettes mortes. Il est alors simple de montrer que, sous le modèle (2), pour tout i, j, k :

$$\log \lambda_{ijk} = s_j + a_{ij} + e_{jk} + p_{ik} \quad (\text{sept})$$

Cette reparamétrisation a de nombreuses propriétés intéressantes: (i) l'équation (7) définit un modèle log-linéaire classique qui pourrait être ajusté en utilisant le maximum de vraisemblance [21]; (ii) les paramètres inconnus a_{ij} , e_{j2} et p_{i2} sont identifiables de manière unique; (iii) le paramètre a_{ij} correspond à la densité relative de la $i^{\text{ème}}$ espèce dans le $j^{\text{ème}}$ DAS mesurée sur une échelle logarithmique, c'est-à-dire qu'il correspond à l'estimation de densité souhaitée, multipliée par une constante inconnue βP_{i1} variable selon les espèces; (iv) le paramètre inconnu e_{j2} correspond à l'effort d'échantillonnage relatif - c'est-à-dire l'effort d'échantillonnage multiplié par une constante inconnue $P_{1k} / (\beta P_{11})$ - des animaux vivants dans le $j^{\text{ème}}$ DAS mesuré sur une échelle logarithmique; (v) le paramètre $e_{j1} = \log(E_{j1} / \beta) = \log V_j$ correspond à l'effort d'échantillonnage connu pour les animaux

morts dans le SAR j (cet effort connu a été inclus dans le modèle comme variable de décalage pour permettre l'ajustement); (vi) les paramètres p_{jk} sont des fonctions compliquées des probabilités de détection. Notons que notre reparamétrisation implique les contraintes d'identifiabilité suivantes: $p_{j1} = p_{j2} = 0$. Ainsi, seuls les paramètres inconnus p_{j2} , $i \neq 1$ ont été estimés.

Prise en compte de la régularité spatiale et du filtrage environnemental

Le modèle développé dans la section précédente a fourni un moyen d'estimer les densités relatives des différentes espèces d'intérêt dans les espèces en péril, tout en tenant compte des inégalités d'effort d'échantillonnage entre les espèces en péril et le statut, ainsi que les différentes détectabilités pour les différentes espèces et le statut. Ainsi, en supposant que l'effort d'échantillonnage était connu dans tous les SAR pour les animaux morts, il aurait pu être possible d'estimer les densités log-relatives $I \times J a_{ij}$, les efforts d'échantillonnage J log-relatifs e_{j2} pour les animaux vivants, ainsi que les paramètres $I - 1$ «bruit» p_{j2} , en maximisant la log-vraisemblance [21]:

$$\log \mathcal{L} = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K N_{ijk} (a_{ij} + e_{jk} + p_{ik}) - \exp(a_{ij} + e_{jk} + p_{ik}) \quad (8)$$

Il s'agit de l'approche classique pour ajuster ce type de modèle log-linéaire, disponible dans tous les logiciels statistiques.

Cependant, à ce stade, nous avons choisi d'introduire des connaissances spatiales et environnementales dans la modélisation par des contraintes sur les paramètres, améliorant ainsi la précision des estimations. D'une part, il était probable que les densités d'espèces étaient spatialement structurées: en raison de la proximité spatiale, deux SAR voisins étaient probablement caractérisés par des profils de densité plus similaires que deux SAR tirés au hasard. D'un autre côté, il était également probable que les densités d'espèces soient affectées par la composition de l'environnement; deux SAR avec un environnement similaire étaient probablement caractérisés par des profils de densité plus similaires que deux SAR avec des conditions très différentes.

Nous avons choisi de rendre compte de cette régularité spatiale et de cet effet environnemental en maximisant une log-vraisemblance régularisée [22, 23]. Nous avons utilisé une méthode de régularisation de type crête, c'est-à-dire une pénalisation de la vraisemblance assurant un certain niveau de régularité spatiale et environnementale, où les différences entre paramètres sont contraintes d'être faibles. Précisément, nous avons considéré la maximisation des critères suivants:

$$\log \mathcal{L} - \sum_{i=1}^I \sum_{j=1}^J \sum_{m=1}^J v \pi_{jm} (a_{ij} - a_{im})^2$$

par rapport aux paramètres a_{ij} , e_{jk} , p_{ik} , où π_{jm} est une mesure des «proximités environnementales et spatiales» entre le SAR j et le SAR m (voir ci-dessous) et v est un paramètre positif qui détermine la force de la pénalité. Il ressort clairement de cette équation que fixer v à une valeur élevée pour une espèce donnée pénaliserait fortement la vraisemblance régularisée si les régions voisines au sens des proximités π_{jm} étaient caractérisées par des estimations de densité fortement différentes. Inversement, régler v à zéro entraînerait l'estimation classique du maximum de vraisemblance.

Nous avons défini les proximités π_{jm} pour prendre en compte les proximités spatiales et environnementales. Nous avons considéré que deux SAR appartenant à la même grande région agricole étaient caractérisées par un environnement plus similaire que deux SAR appartenant à différentes grandes régions agricoles. Nous avons défini les proximités spatiales et environnementales de la manière suivante: (i) $\pi_{jm} = 1$ lorsque les SAR j et m sont spatialement adjacents et appartiennent à la même grande région agricole, (ii) $\pi_{jm} = 0,5$ soit lorsque les SAR j et m sont spatialement adjacents mais n'appartiennent pas à la même grande région agricole (le cas le plus courant où $\pi_{jm} = 0,5$), ou ne sont pas spatialement adjacents mais appartiennent à la même grande région agricole (ce qui est plus rare, car 83% des grandes régions agricoles contiennent au plus deux SAR), et (iii) $\pi_{jm} = 0$ sinon. Notez qu'une approche de validation croisée a indiqué qu'une régularisation basée sur des proximités tenant compte à la fois de la structure spatiale et environnementale a entraîné une erreur de prédiction beaucoup plus petite qu'une régularisation basée sur des proximités purement spatiales, suggérant ainsi que notre combinaison d'espace et d'environnement était nécessaire pour la estimation de la densité (voir l'annexe S1, section 7).

Nous avons sélectionné la meilleure valeur du paramètre de retrait v avec une approche de validation croisée R -fold (p. 241 dans [22]: nous avons divisé les données en R parties, chaque partie correspondant à une année de collecte de données (ie $R =$ Pour la $r^{\text{ème}}$ partie et pour une valeur donnée de v , nous avons ajusté le modèle aux autres $R - 1$ parties des données et calculé l'erreur de prédiction du modèle ajusté lors de la prédiction de la $r^{\text{ème}}$ partie des données, en utilisant la log-vraisemblance négative de la partie prédite des données, c'est-à-dire:

$$Q^{(r)}(v) = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K N_{ijk}^{(r)} \left\{ s_j + a_{ij}^{(-r)} + e_{jk}^{(-r)} + p_{ik}^{(-r)} - \log(R - 1) \right\} - \exp \left\{ s_j + a_{ij}^{(-r)} + e_{jk}^{(-r)} + p_{ik}^{(-r)} - \log(R - 1) \right\} \quad (9)$$

où l'exposant $(-r)$ indique que le paramètre a été estimé avec la $r^{\text{ème}}$ partie des données supprimée, et $N_{jejk}^{(r)}$ est le nombre d'animaux de l'espèce i détectés avec le statut k dans la $j^{\text{ème}}$ RAS au cours de la $r^{\text{ème}}$ année. Notez que le log des composants ($R - 1$) apparaissant deux fois dans la formule ci-dessus permet la prédiction du nombre $N_{jejk}^{(r)}$ d'animaux détectés sur une seule année grâce à l'effort $e_{jk}^{(-r)}$ estimé sur $R - 1$ ans. Nous avons calculé $Q^{(r)}$ pour $r = 1, 2, \dots, R$ et combiné les estimations R de l'erreur de prédiction avec:

$$Q(v) = \sum_{r=1}^R Q^{(r)}(v) \quad (\text{dix})$$

Nous avons sélectionné la valeur de v pour laquelle l'erreur de prédiction $Q(v)$ était la plus petite (notez que d'autres critères mesurant l'erreur de prédiction ont donné des résultats similaires; voir l' [annexe S1](#) , section 4.1.1).

Deux mesures d'effort possibles pour les animaux morts

Nous ne savions pas exactement le nombre de kilomètres parcourus par les officiers dans chaque SAR au cours de la période d'étude. Cependant, la base de données SCSL contenait suffisamment d'informations pour l'estimer. Nous avons comparé deux façons différentes d'estimer le nombre de kilomètres parcourus par les agents dans chaque SAR au cours de la période d'étude (voir également la section 2.3 de l' [annexe S1](#)).

Une première mesure $V_j^{(1)}$ pourrait être dérivée du nombre total de kilomètres parcourus dans un département (calculé avec l' [équation 1](#)) en supposant que l'effort était uniformément réparti dans le département. Selon cette hypothèse, nous pourrions estimer le nombre de kilomètres parcourus dans chaque SAR en redistribuant ce nombre total entre les SAR en fonction de leur superficie. Ainsi, S_j étant la zone couverte par le $j^{\text{ème}}$ SAR, nous avons approché le nombre de kilomètres $V_j^{(1)}$ voyagé dans chaque SAR avec:

$$V_j^{(1)} = T_d \times \frac{S_j}{\sum_{j \in D} S_j}$$

où D est l'ensemble des SAR appartenant au département d .

Nous avons également envisagé une autre mesure $V_j^{(2)}$, généré par un modèle plus compliqué attribuant le nombre total de kilomètres parcourus par les agents d'un département aux SAR de ce département: ce modèle alternatif rendait compte du fait qu'une voiture donnée concentre généralement son activité dans une zone restreinte du département (chaque brigade travaille généralement sur un «territoire» informel). Pour chaque voiture, nous avons calculé un barycentre d'activité en fonction de la localisation moyenne des animaux détectés par cette voiture. Regroupées sur toutes les voitures du département, la répartition des distances entre les animaux détectés et ces barycentres était log-normale avec une moyenne et un écart type selon le département. Par conséquent, pour chaque département et chaque voiture en particulier, nous avons pu estimer une fonction bidimensionnelle donnant la densité de probabilité de présence de la voiture en chaque point. En faisant la moyenne de ces densités sur l'ensemble des voitures, nous avons obtenu une fonction donnant la densité de probabilité de présence d'une voiture en chaque point du département. Après intégration de cette fonction de densité de probabilité sur un SAR donné au sein d'un département, on a pu estimer la proportion de temps passé par les voitures du département dans le SAR. On pourrait enfin estimer le nombre de kilomètres parcourus par les voitures d'un département dans un SAR en multipliant le nombre total de kilomètres on pourrait estimer la proportion de temps passé par les voitures du département dans la RAS. On pourrait enfin estimer le nombre de kilomètres parcourus par les voitures d'un département dans un SAR en multipliant le nombre total de kilomètres on pourrait estimer la proportion de temps passé par les voitures du département dans la RAS. On pourrait enfin estimer le nombre de kilomètres parcourus par les voitures d'un département dans un SAR en multipliant le nombre total de kilomètres T_d avec cette proportion.

Pour identifier la meilleure mesure de l'effort, nous avons comparé l'efficacité prédictive de notre modèle lorsque l'effort d'échantillonnage des morts était mesuré soit avec $V_j^{(1)}$ ou avec $V_j^{(2)}$. Nous avons utilisé une approche de validation croisée: nous avons considéré séparément les années paires (2002, 2004) et les années impaires (2003, 2005) de notre période d'étude; pour chaque mesure, nous avons ajusté un modèle en utilisant les données collectées pendant les années paires (en utilisant pour chacune la valeur optimale de v , choisie selon l'approche décrite dans la section précédente) et nous avons mesuré l'erreur de prédiction sur les années impaires, et inversement. L'erreur de prédiction a été mesurée à l'aide du critère Q décrit dans la section précédente (voir section 6 de l' [annexe S1](#)).

Le modèle a été ajusté à l'aide d'un programme C en interface avec le logiciel R [24]. Ce programme est disponible dans la fonction `penalizedmodel()` du package `scsl` ([S1 Package](#)). De plus, nous avons bootstrapé nos données 100 fois pour permettre le calcul des coefficients de variation pour les estimations de densité relative.

Déclaration d'éthique

Le travail de terrain a été effectué par 1500 agents professionnels travaillant pour l'ONCFS. La collecte de données pour le programme SCSL faisait partie de leur travail, de sorte que leur consentement écrit ou verbal pour cette étude spécifique n'était pas nécessaire. Ces officiers qualifiés étaient mandatés par le ministre français de l'Écologie, ce qui leur permettait de travailler avec des animaux morts, conformément à la législation française. Le programme SCSL a été approuvé par le directeur de l'ONCFS.

Résultats

Nous avons ajusté le modèle log-linéaire pénalisé développé dans la section précédente sur l'ensemble de données complet, y compris les données concernant la genette commune et le chat sauvage. En effet, même si ces deux espèces ne nous intéressaient pas, l'inclusion de ces données dans l'ajustement ajoutait de la précision à l'estimation de l'effort e_{j2} (ces détections prouvaient en fait la présence des officiers à ces endroits).

Nous avons d'abord comparé les deux mesures possibles de l'effort d'échantillonnage en utilisant la validation croisée. La log-vraisemblance négative de l'ensemble de données estimée avec la validation croisée était beaucoup plus élevée pour la mesure $V_j^{(2)}$ ($Q_2 = 5468$) que pour la mesure $V_j^{(1)}$ ($Q_1 = 5240$), indiquant que ce dernier était caractérisé par la meilleure efficacité prédictive (voir l' [annexe S1](#) , section 6, pour plus de détails). Ce résultat peut s'expliquer par le fait que le rôle principal des agents nationaux de protection de la faune est de veiller à ce que l'application des lois relatives à la chasse et au piégeage des animaux sauvages soit la plus uniforme possible sur un département, ce qui suggère que leur fréquence d'utilisation des SAR pendant la période d'étude était proportionnelle à leur superficie. Nous avons donc ajusté notre modèle en utilisant la mesure $V_j^{(1)}$ de l'effort d'échantillonnage.

Pour ce modèle, la valeur de $\nu = 0,3$ était le meilleur choix pour minimiser l'erreur de prédiction pour toutes les espèces (Fig. 2). Cette petite pénalité indique qu'à cette échelle, les modèles spatiaux et environnementaux d'abondance ne sont pas très forts pour la plupart des espèces (Fig. 3). L'examen des résidus n'a révélé aucune tendance problématique et la qualité de l'ajustement était satisfaisante. Nous avons également vérifié l'absence de surdispersion dans notre ajustement en (i) vérifiant que les résidus normalisés étaient caractérisés par une variance constante lorsqu'ils étaient tracés en fonction des valeurs prédites, et (ii) calculant le paramètre de dispersion associé à ce modèle. Cette statistique de Pearson était égale à $\chi^2 = 5962$, pour un nombre de degrés de liberté égal à $d = 4913$: le paramètre de dispersion a donc été estimé à $\hat{h} = \chi^2 / j = 1$ (p. 175 dans [25]), une valeur très proche de 1 (voir [26] pour une discussion sur ce paramètre de dispersion et l'identification de la surdispersion). L'absence de surdispersion dans notre modèle était donc une hypothèse raisonnable (voir la section 8 de l' annexe S1 pour plus de détails).

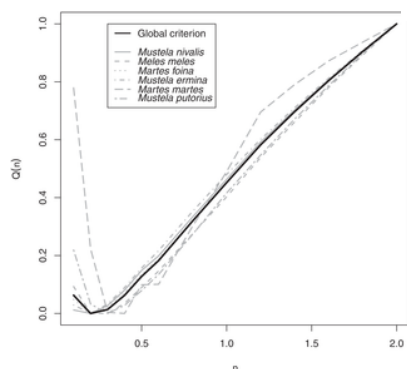


Fig 2. Valeur de l'erreur de prédiction $Q(\nu)$ pour différentes valeurs du paramètre de pénalité ν calculé par l'approche de validation croisée. Nous montrons également les contributions $Q^i(\nu)$ de chaque espèce i à cette erreur de prédiction. Pour faciliter l'interprétation visuelle de ces résultats, les critères $Q(\nu)$ et $Q^i(\nu)$ ont été remis à l'échelle entre 0 et 1 dans la plage de ν d'intérêt.

<https://doi.org/10.1371/journal.pone.0121689.g002>

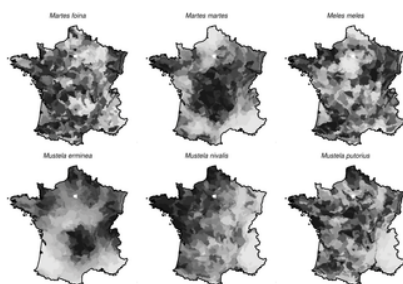


Fig 3. Densité relative des 6 espèces de *Mustelidae* d'intérêt telle qu'estimée par notre modèle (les zones plus sombres correspondent à des densités plus élevées).

<https://doi.org/10.1371/journal.pone.0121689.g003>

Les densités estimées (Fig. 3) pour chaque espèce (animaux morts et vivants) ont confirmé les modèles connus de distribution, mais ont amélioré notre compréhension à une plus petite échelle. La belette est apparue plus abondante dans le nord-ouest de la France, et particulièrement en Bretagne. L'hermine était plus abondante dans le Massif Central, la région montagneuse du centre sud de la France, ainsi que dans l'est et le nord de la France. Le blaireau et la martre d'Amérique étaient présents dans tout le pays, avec des densités probablement variables localement. Le putois était rare dans les Alpes et était plus fréquent dans les régions les plus humides du pays (par exemple, en Bretagne). La martre des pins était abondante dans le Massif Central et en Bretagne.

Nous avons également estimé les coefficients de variation associés aux estimations de densité relative pour chaque espèce et chaque SAR (Fig. 4). En général, le coefficient de variation de l'estimation de la densité relative dans un SAR était compris entre 15% et 35%.

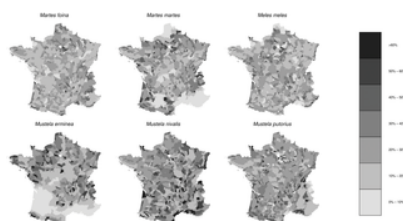


Fig 4. Coefficients de variation correspondant aux estimations de densité relative pour les 6 espèces de *Mustelidae* d'intérêt.

<https://doi.org/10.1371/journal.pone.0121689.g004>

Discussion

Jusqu'à présent, les données collectées dans le cadre du programme SCSL n'étaient utilisées que par les décideurs et les sociétés de conseil privées pour établir des cartes de distribution de l'espèce à partir des nombres bruts d'animaux détectés. Cependant, comme de nombreux auteurs l'ont déjà noté, ne pas tenir compte des efforts d'échantillonnage inégaux ainsi que des probabilités de détection inégales empêche toute estimation de la densité [18]. Nous avons donc développé un modèle prenant en compte un effort d'observation inégal selon les régions et l'état de détection (animal mort / vivant), ainsi qu'une probabilité de détection variable selon les espèces et le statut, pour estimer la densité relative de 6 *Mustelidae* espèces dans chaque petite région agricole de France. La prise en compte de cet effort dans notre modèle a conduit à des différences substantielles entre les densités estimées et les nombres bruts d'animaux détectés (voir Fig. 5).

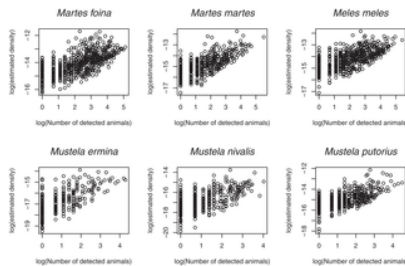


Fig 5. Relation entre la densité relative estimée par notre modèle et le nombre brut de détection dans chaque petite région agricole (SAR).

Ces variables sont présentées sur une échelle logarithmique (nous nous sommes concentrés sur le DAS pour lequel la densité estimée et le nombre brut de détections étaient supérieurs à 0).

<https://doi.org/10.1371/journal.pone.0121689.g005>

Notre modèle suppose l'absence d'erreurs dans l'identification des espèces. De telles erreurs peuvent entraîner une surestimation de l'abondance d'une espèce rare si elle est confondue avec une espèce commune [27]. Dans cette étude, les observateurs étaient des professionnels formés pour collecter des données pour des programmes scientifiques; lorsqu'ils n'étaient pas certains de l'espèce détectée, ils n'ont jamais signalé la détection. Ainsi, dans une certaine mesure, l'incertitude d'identification a conduit à une probabilité de détection plus faible (voir ci-dessous). Cependant, le fait qu'un agent était certain de l'identification de l'espèce ne signifie pas qu'il a effectivement identifié l'espèce correctement. Une erreur d'identification était possible, surtout lorsqu'un animal vivant était détecté, étant donné qu'un examen attentif de l'animal n'était en général pas possible dans ce cas. Les erreurs d'identification étaient moins probables lorsqu'un animal mort était détecté, ce qui était le cas de 74% des blaireaux, putois et martres. De plus, les animaux vivants étaient souvent détectés la nuit, alors que l'identification visuelle était encore plus difficile. De telles erreurs étaient également susceptibles de varier selon les agents, surtout avec l'expérience. Enfin, les erreurs d'identification étaient susceptibles de dépendre de l'espèce: elles devaient être rares pour les blaireaux bien identifiables. D'un autre côté, la martre des pins et la martre des pins sont caractérisées par des apparences et des tailles similaires et étaient plus susceptibles d'être confondues, bien que les différences d'habitat utilisé par les deux espèces [28 - 30] —connu des officiers - peut avoir limité ces confusions. L'hermine et la belette pourraient également avoir été confondues, mais dans une moindre mesure depuis (i) ces espèces ont été détectées vivantes pendant la journée, (ii) elles sont caractérisées par des tailles différentes, et (iii) la pointe noire prédominante de la queue de l'hermine est très visible pendant la journée. Des études supplémentaires sont nécessaires pour tenir compte de ce type d'erreur dans notre modèle. Une approche possible serait de modéliser la probabilité d'erreur d'identification à l'aide d'un ensemble de données collectées par un échantillon aléatoire d'agents de l'ONCFS avec une approche à double observateur [27]. Nous pourrions alors adapter notre approche dans un cadre bayésien, en tenant compte de l'incertitude causée par les erreurs d'identification à l'aide d'une distribution a priori informative dérivée de ce modèle d'erreurs d'identification.

Dans cette étude, nous avons supposé que l'effort d'échantillonnage des animaux morts était proportionnel au nombre de kilomètres parcourus par les agents. Cependant, il s'agit d'une approximation très approximative de l'effort réel, et il serait intéressant de modéliser cet effort plus précisément dans une étude ultérieure. En effet, cet effort d'échantillonnage peut être vu comme la probabilité qu'un animal donné a dans une région donnée ait été détecté comme un animal mort par un agent. Si nous simplifions le système en considérant que tous les animaux morts détectés sont morts à la suite d'un massacre sur la route, cette probabilité est égale à la probabilité $p_0(a)$ qu'un animal donné a été tué sur la route, multipliée par la probabilité $p_1(a)$ que cet animal tué a été détecté et signalé par les agents. Notre modèle ne tient pas compte de tous les facteurs affectant ces deux probabilités, ce qui peut avoir donné lieu à des estimations biaisées. D'une part, la probabilité $p_0(a)$ n'était probablement pas constante dans toutes les ZA: en effet, la mortalité routière est plus fréquente non seulement là où l'espèce est abondante, mais aussi là où la charge de trafic est importante [31, 32]. Par conséquent, toutes choses étant égales par ailleurs, les SAR avec une charge de trafic plus importante étaient probablement caractérisés par un plus grand nombre de tués sur la route. Ne pas tenir compte de la différence de charge de trafic dans notre modèle peut donc avoir entraîné une surestimation de l'abondance relative dans les SAR plus urbaines. Par contre, cet effet a été atténué par le fait que, dans une RS donnée, les agents de protection de la faune passaient leur temps là où se trouvait la faune, c'est-à-dire en milieu rural, à rouler sur des routes à trafic léger à modéré. Ainsi, au sein d'un SAR, $p_1(a)$ était inversement proportionnel à $p_0(a)$, limitant ce biais. D'autres facteurs peuvent également avoir affecté $p_0(a)$, comme les courbes sur les routes et les perturbations humaines [33], mais la manière dont ces facteurs à petite échelle ont affecté la probabilité globale $p_0(a)$ dans un SAR donné n'est pas encore claire, et des recherches supplémentaires sont nécessaires pour modéliser plus précisément l'échantillonnage l'effort des animaux morts en fonction de toutes les variables affectant significativement $p_0(a)$ et $p_1(a)$.

Notre approche suppose également que la densité d'une espèce le long des routes d'une RS était proportionnelle à sa densité dans les endroits où les agents passaient leur temps lors d'activités statiques. Étant donné que les activités statiques devaient être réparties uniformément sur les zones où la faune est présente (c'est-à-dire sur les types d'habitats non urbains), cette hypothèse aurait pu être violée de deux façons: (i) à grande échelle, si les proportions d'habitat les types disponibles dans une RS donnée ne correspondaient pas aux proportions de ces types d'habitats le long des routes, et (ii) à petite échelle, si la densité locale des espèces le long d'une route était différente de la densité d'intérêt dans l'habitat qui il a traversé. Le point (i) était improbable, car les routes sont bien réparties sur le territoire français, à forte densité (1,77 km / km², étant en moyenne comparable à la densité des routes au Royaume-Uni, en Italie ou en Suisse), et parce qu'il n'y a aucune raison pour que les routes évitent certains types d'habitats. Concernant le point (ii), nous notons que les routes ne constituaient pas une barrière aux déplacements

des petits animaux carnivores. Il a en effet été démontré que ces animaux traversent les routes dans des études de radio-pistage (Ruelle, com. Pers.). Bien entendu, le fait que les routes ne constituent pas une barrière ne signifie pas qu'elles ont été utilisées avec la même fréquence que l'habitat environnant. Cependant, s'il y avait un biais constant (par exemple si la densité d'une espèce donnée le long des routes était toujours 20% inférieure à la densité dans l'habitat environnant), cette différence pourrait être considérée comme faisant partie de la détectabilité imparfaite des animaux dans notre modèle et n'entraînerait pas d'estimations biaisées. En revanche, si le biais n'était pas constant (par exemple, si la probabilité qu'un animal traverse la route était plus élevée lorsque la route était située dans une forêt), cela affecterait en fait les estimations. Des recherches supplémentaires sont nécessaires pour étudier comment la sélection de l'habitat par ces petites espèces carnivores varie en fonction de la distance aux routes.

Il est bien connu que les capacités de détection de l'observateur peuvent varier et avoir des conséquences problématiques dans les études sur l'abondance des espèces [34]. En effet, pour un nombre donné de kilomètres parcourus, un agent donné n'a observé la route et les bords de route que sur une fraction de ces kilomètres. Cette fraction était probablement variable parmi les officiers. Dans notre contexte de modélisation, une probabilité de détection variant d'un observateur à l'autre était mathématiquement identique à un effort d'échantillonnage variant d'un observateur à l'autre. Puisqu'une équipe d'agents donnée a concentré son activité sur une zone réglementée au sein d'un ministère, il s'ensuit que le nombre de kilomètres parcourus par l'équipe dans cette zone n'était pas nécessairement exactement proportionnel à son effort d'échantillonnage. Pour un nombre donné de kilomètres parcourus dans une région, une équipe avec une probabilité de détection plus élevée était caractérisée par un «effort» plus important qu'une équipe avec une probabilité plus faible. Cependant, nous avons mesuré l'effort d'échantillonnage des animaux morts en additionnant les efforts individuels estimés, c'est-à-dire une opération mathématique identique à la moyenne de ces efforts, jusqu'à une constante. Comme il y avait en moyenne 9 équipes travaillant sur chaque SAR (intervalle interquartile: 5-13) et que chaque équipe était généralement composée de deux officiers, la moyenne de l'effort d'échantillonnage de toutes les équipes travaillant dans un SAR donné a fortement diminué l'hétérogénéité des compétences de détection. De plus, si l'on suppose que les capacités de détection et la motivation des agents ne sont pas des variables spatialement autocorrélées, l'effort d'échantillonnage moyen devrait être proportionnel au nombre total de kilomètres dans une région donnée. Compte tenu de l'homogénéité du réseau des agents nationaux de protection de la faune en termes de tâches et de formation, une autocorrélation spatiale de leurs compétences n'était pas attendue. cependant, nous ne pouvons garantir l'absence de motivation spatialement structurée des agents (c'est-à-dire que les agents auraient pu être plus motivés à signaler une détection à certains endroits qu'à d'autres). Il serait donc intéressant d'étudier cette hétérogénéité à partir d'une étude spécifique, pour apprécier le biais provoqué par cette variation dans l'espace et dans le temps de la motivation.

Nous avons utilisé une régularisation de type crête pour améliorer la prédiction de notre modèle en tenant compte d'une structure spatiale et environnementale dans la densité des espèces. Silverman [35] a noté ce qui suit à propos d'une autre méthode statistique reposant sur un paramètre de lissage: «le processus d'examen de plusieurs tracés des données, tous lissés par des quantités différentes, peut bien donner plus d'informations sur les données que de simplement considérer une seule courbe produite automatiquement ». Bien que ce ne soit pas l'objectif principal de notre étude, cette remarque s'applique également à notre approche de modélisation. En effet, nous avons remplacé les métriques de proximité spatiale et environnementale utilisées dans notre régularisation par une proximité purement spatiale, avec $\pi_{jm} = 1$ lorsque les SAR j et m sont adjacents, et $\pi_{jm} = 0$ sinon. La définition du paramètre ν à une valeur très élevée ($\nu = 20$) nous a permis d'identifier les modèles de distribution à grande échelle de l'espèce sur la zone d'étude (Fig. 6). Bien que ces cartes soient caractérisées par un biais plus important et ne devraient donc pas être utilisées pour la gestion, elles peuvent être utiles à des fins d'exploration. Ces cartes illustrent en effet plus clairement les vastes zones de présence identifiées sur la figure 3 .

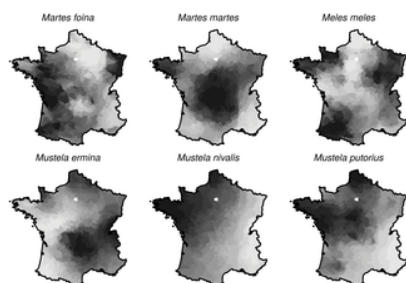


Fig 6. Distribution spatiale des 6 espèces de *Mustelidae* d'intérêt telles qu'estimées par un modèle purement régularisé spatialement, avec un paramètre de pénalité élevé $\nu = 20$ (les zones plus sombres correspondent à des densités plus élevées).

<https://doi.org/10.1371/journal.pone.0121689.g006>

Dans cet article, nous avons modélisé la densité relative des *Mustelidae* espèces à partir de données collectées sur 4 ans. Comme pour toutes les études visant l'estimation de l'abondance, la densité d'une espèce donnée peut avoir changé au cours de notre étude. Étant donné que le nombre d'agents travaillant pour le programme varie peu d'une année à l'autre et que cette période est courte, il est acceptable de supposer un effort constant au sein d'une RS donnée pendant la période d'étude. Sous cette hypothèse, le modèle nous a en effet permis d'estimer la densité relative moyenne de chaque espèce dans chaque SAR, qui était la quantité d'intérêt pour nous. Cependant, même si l'étude de ces changements ne nous intéresse pas, il aurait été possible de les étudier avec notre cadre de modélisation, simplement en remplaçant le facteur «SAR» dans notre modèle par un autre facteur combinant les SAR et les années en l'ajustement, menant à des estimations de la densité relative de l'espèce dans chaque ZPE au cours de chaque année. Notez que nous fournissons à la fois les données et les fonctions requises pour cet ajustement dans le package compagnon `R_scs1` (voir [Package S1](#)), afin que le lecteur intéressé puisse essayer cette approche «spatio-temporelle» sur nos données. A noter que si un suivi de la densité relative dans le temps est effectué avec cette approche, il est important d'éviter une diminution de la motivation des agents avec le temps (qui conduirait à une diminution non mesurée de l'effort d'échantillonnage, c'est-à-dire non strictement liée à le nombre de kilomètres). Dans notre étude, les données collectées via le programme `SCSL` sont des données très simples et rapides à collecter. Ils sont utiles aux agents eux-mêmes à l'échelle locale (par exemple pour construire des cartes régionales de distribution), ce qui les aide beaucoup à les garder motivés. Nous avons également essayé de motiver les gens grâce à des rapports annuels et des communications régulières, ce qui les aide à rester intéressés et informés du projet.

Enfin, notre approche de modélisation a en elle-même des implications intéressantes. En effet, cette approche permet de combiner deux ensembles de données collectées sur la même population: (i) un ensemble de données se concentrant sur les animaux morts collectés avec un effort d'échantillonnage supposé connu, et (ii) un ensemble de données collectées sur les animaux vivants avec un échantillonnage inconnu effort. Cette capacité de notre approche de modélisation à combiner des informations collectées avec différentes méthodes sur une même population a des implications intéressantes pour l'analyse des données collectées dans les programmes de sciences citoyennes [18]. En effet, notre cadre de modélisation pourrait en effet être utilisé pour combiner (i) un jeu de données collecté à l'aide d'un plan d'échantillonnage connu généralement conçu par une équipe scientifique et (ii) un jeu de données opportuniste obtenu par crowdsourcing, collecté par des volontaires, généralement sans aucune connaissance de l'échantillonnage. effort. Notre approche nous permettrait d'estimer l'effort d'échantillonnage pour l'ensemble de données opportuniste et de corriger les biais générés par l'effort incontrôlé et inconnu dans ces données. Les résultats préliminaires suggèrent que cette approche nous permettrait d'obtenir des estimations non biaisées et précises des densités relatives des espèces étudiées (Giraud et al. En préparation). Cela confirme la remarque de MacKenzie et al. [36]: «Dans certaines situations, il peut être approprié de partager ou d'emprunter des informations sur les paramètres de population d'espèces rares à partir de plusieurs sources de données. Le concept général est qu'en combinant les données, le cas échéant, des estimations plus précises des paramètres peuvent être obtenues. » Des travaux supplémentaires sont désormais nécessaires pour identifier les conditions dans lesquelles notre approche pourrait être utilisée avec les données collectées dans le cadre des programmes de science citoyenne.

Renseignements à l'appui

Annexe S1. Détails des calculs effectués pour ajuster les modèles.

<https://doi.org/10.1371/journal.pone.0121689.s001>

(PDF)

Paquet S1. Package R contenant les jeux de données et les fonctions.

<https://doi.org/10.1371/journal.pone.0121689.s002>

(GZ)

Remerciements

Nous remercions chaleureusement les membres du groupe CiSStats pour des discussions stimulantes et fructueuses sur les données opportunistes et les questions statistiques connexes. Nous remercions également Nigel Gilles Yoccoz et un arbitre anonyme pour leurs commentaires très utiles, qui ont grandement contribué à améliorer cet article. Ces travaux ont été partiellement financés par le programme Mastodons du CNRS, par le programme CiSStats de l'INRA et par la Chaire de Modélisation Mathématiques et Biodiversité de VEOLIA-Ecole Polytechnique-MNHN. Nous remercions également Murielle Guinot-Ghestem (ONCFS) pour ses commentaires sur les versions précédentes du manuscrit, et American Journal Experts pour la vérification de la langue anglaise.

Les contributions de l'auteur

Conception et réalisation des expériences: CC JC CG SH RJ PM JP DP SR. Réaliser les expériences: CC JC CG SH RJ PM JP DP SR. Analyse des données: CC JC CG SH RJ PM JP DP SR. Réactifs / matériaux / outils d'analyse fournis: CC JC CG SH RJ PM JP DP SR. Rédaction de l'article: CC JC CG SH RJ PM JP DP SR.

Références

1. MacKenzie DI, Kendall WL (2002) Comment intégrer la probabilité de détection aux estimations de l'abondance relative? *Ecology* 83: 2387-2393.
[Voir l'article](#) • [Google Scholar](#)
2. Johnson D (2008) En défense des indices: le cas des relevés d'oiseaux. *J Wildl Manage* 72: 857-868.
[Voir l'article](#) • [Google Scholar](#)
3. Seber G (1982) L'estimation de l'abondance animale et des paramètres connexes. Charles Griffin & compagnie Ltée.
4. Buckland S, Anderson D, Burnham K, Laake J (1993) *Distance Sampling: Estimating Abundance of Biological Populations*. Chapman et Hall.
5. Kindberg J, Ericsson G, Swenson JE (2009) Surveillance de grands mammifères rares ou insaisissables à l'aide d'observateurs volontaires à e ort corrigé. *Biol Conserv* 142: 159-165.
[Voir l'article](#) • [Google Scholar](#)
6. Gibbons DW, Donald PF, Bauer HG, Fornasari L, Dawson IK (2007) Cartographie des distributions aviaires: l'évolution des atlas d'oiseaux. *Étude d'oiseau* 54: 324-334.
[Voir l'article](#) • [Google Scholar](#)
- sept. Kurki S, Nikula A, Helle P, Linden H (1998) Abondances de renard roux et de martre des pins par rapport à la composition des paysages de la forêt boréale. *Journal de l'écologie animale* 67: 874-886.
[Voir l'article](#) • [Google Scholar](#)
8. Birks J, Messenger J, Braithwaite T, Davison A, Brookes R, Strachan C (2005) Les relevés d'excréments sont-ils une méthode fiable pour évaluer la répartition et l'état de la population de martres des pins? Dans: Harrison D, Fuller A, Proulx G, éditeurs, *Martens and Fishers in Human-Altered Environments*, Springer-Verlag, Berlin. 235-252.
9. Livia L, Vercillo F, Palomba A, Panara F, Ragni B (2007) Une méthode pcr-rflp sur des échantillons fécaux pour distinguer *Martes martes*, *Martes foina*, *Mustela putorius* et *Vulpes vulpes*. *Conservation Genetics* 8: 757-759.
[Voir l'article](#) • [Google Scholar](#)
- dix. Harris S, Yalden D (2004) Un programme de surveillance intégré pour les mammifères terrestres en Grande-Bretagne. *Mammal Review* 34: 157-167.
[Voir l'article](#) • [Google Scholar](#)

11. Davison A, Birks JD, Brookes RC, Braithwaite TC, Messenger JE (2002) Sur l'origine des fèces: méthodes morphologiques versus moléculaires pour surveiller les carnivores rares à partir de leurs excréments. *Journal of Zoology* 257: 141-143.
[Voir l'article](#) • [Google Scholar](#)
12. Wilson GJ, Delahay RJ (2001) Un examen des méthodes d'estimation de l'abondance des carnivores terrestres à l'aide de signes de terrain et d'observation. *Wildlife Research* 28: 151-164.
[Voir l'article](#) • [Google Scholar](#)
13. Hounsom T, Young R, Davison J, Yarnell R, Trewby I, Garnett BT, et al. (2005) Une évaluation de l'échantillonnage à distance pour estimer l'abondance du blaireau (*Meles meles*). *Journal of Zoology* 266: 81-87.
[Voir l'article](#) • [Google Scholar](#)
14. Ruetter S, Stahl P, Albaret M (2003) Facteurs affectant le succès de piégeage des renards, martres et martres des pins. *Biologie de la faune* 9: 11-19.
[Voir l'article](#) • [Google Scholar](#)
15. Albaret M, Ruetter S (2008) Enquête nationale sur les pièges à l'usage de sept espèces de martres au cours de la saison 2007-2008. *Faune Sauvage* 295: 39-43.
[Voir l'article](#) • [Google Scholar](#)
16. Sadler LM, Webbon CC, Baker PJ, Harris S (2004) Méthodes de surveillance des renards roux *Vulpes vulpes* et blaireaux *Meles meles*: les signes de terrain sont-ils la réponse? *Mammal Review* 34: 75-98.
[Voir l'article](#) • [Google Scholar](#)
17. Barea-Azcárate JM, Virgós E, Ballesteros-Duperón E, Mole M, Chiroso M (2007) Arpentage des niveaux de voiture à de grandes échelles spatiales: une comparaison de quatre larges méthodes appliquées. *Biodiversité et conservation* 16: 1213-1230.
[Voir l'article](#) • [Google Scholar](#)
18. Hochachka WM, Fink D, Hutchinson RA, Sheldon D, Wong WK, Kelling S (2012) Science intensive en données appliquée à la science citoyenne à grande échelle. *Trends Ecol Evol* 27: 130-137. pmid: 22192976
[Voir l'article](#) • [PubMed / NCBI](#) • [Google Scholar](#)
19. Xie Y (2013) knitr: Un package à usage général pour la génération de rapports dynamiques dans R. Version du package R 12.
20. Darroch J, Speed T (1983) Modèles et interactions additifs et multiplicatifs. *Ann Stat* 11: 724-738.
[Voir l'article](#) • [Google Scholar](#)
21. Cameron A, Trivedi P (1998) Analyse de régression des données de dénombrement. *Monographies de la société économétrique*, n° 30.
22. Hastie T, Tibshirani R, Friedman J (2009) Les éléments de l'apprentissage statistique. Exploration de données, inférence et prédiction. Deuxième édition. Série Springer en statistiques.
23. Malbasa V, Vucetic S (2011) Régression logistique spatialement régularisée pour la cartographie des maladies sur de grandes populations mobiles. In: Actes de la 17e conférence internationale ACM SIGKDD sur la découverte de connaissances et l'exploration de données. ACM, pp. 1352-1360.
24. R Core Team (2013) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienne, Autriche. URL <http://www.R-project.org/>.
25. McCullagh P, Nelder J (1989) Modèles linéaires généralisés. Deuxième édition. Londres: Chapman & Hall.
26. Lindsey J (1999) Sur l'utilisation des corrections pour la surdispersion. *Journal de la Royal Statistical Society: Série C (Statistiques appliquées)* 48: 553-561.
[Voir l'article](#) • [Google Scholar](#)
27. Conn PB, McClintock BT, Cameron MF, Johnson DS, Moreland EE, Boveng PL (2013) Accommodant les erreurs d'identification des espèces dans les relevés de transect. *Ecology* 94: 2607-2618. pmid: 24400512
[Voir l'article](#) • [PubMed / NCBI](#) • [Google Scholar](#)
28. Rondinini C, Boitani L (2002) Utilisation de l'habitat par la martre du hêtre dans un paysage fragmenté. *Ecographie* 25: 257-264.
[Voir l'article](#) • [Google Scholar](#)
29. Zalewski A, Jedrzejewski W (2006) Organisation spatiale et dynamique de la population de martre des pins dans la forêt de Biały (Pologne) par rapport à d'autres forêts européennes. *Ecography* 29: 31-43.
[Voir l'article](#) • [Google Scholar](#)
30. Mergery M, Helder R, Roeder JJ (2011) Effet de la fragmentation des forêts sur les modèles d'utilisation de l'espace chez la martre des pins (martre des pins). *Journal of Mammalogy* 92: 328-335.
[Voir l'article](#) • [Google Scholar](#)
31. Baker PJ, Harris S, Robertson CP, Saunders G, White PC (2004) Est-il possible de surveiller les changements de population de mammifères à partir des dénombrements des accidents de la route? une analyse utilisant le renard roux de Bristol *Vulpes vulpes* comme exemple. *Mammal Review* 34: 115-130.
[Voir l'article](#) • [Google Scholar](#)

32. Clarke G, White P, Harris S (1998) Effets des routes sur le blaireau meles meles populations dans le sud-ouest de l'Angleterre. *Biological Conservation* 86: 117-124.
[Voir l'article](#) • [Google Scholar](#)
33. Grilo C, Ascens [notdef] o F, Santos-Reis M, Bissonette JA (2011) Les paysages bien connectés favorisent-ils la mortalité routière? *Journal européen de recherche sur la faune* 57: 707–716.
[Voir l'article](#) • [Google Scholar](#)
34. Link W, Sauer J (2002) Une analyse hiérarchique du changement de population avec application à la paruline azurée. *Ecology* 83: 2832-2840.
[Voir l'article](#) • [Google Scholar](#)
35. Silverman B (1986) Estimation de la densité pour les statistiques et l'analyse des données. Londres: Chapman et Hall.
36. MacKenzie DI, Nichols JD, Sutton N, Kawanishi K, Bailey LL (2005) Amélioration des inférences dans les études de population d'espèces rares détectées imparfaitement. *Ecology* 86: 1101-1113.
[Voir l'article](#) • [Google Scholar](#)