

Déterminants des collisions routières de la faune sauvage : modélisation spatiale intégrée de données protocolées et opportunistes

MaëliSS VANDROUX

Stage encadré par Jean-Yves BARNAGAUD, Jean-François BRETAUD et Alain MORAND

Centre d'Études et d'expertise sur les Risques, l'Environnement, la Mobilité et l'Aménagement

Centre d'Écologie Fonctionnelle et Évolutive

Université Toulouse III Paul Sabatier

Rapport réalisé dans le cadre d'un stage de master 2

Parcours Modélisation des Systèmes Ecologiques



École Pratique
des Hautes Études

PSL



AGIR pour la
BIODIVERSITÉ
BRETAGNE



AGIR pour la
BIODIVERSITÉ
LOIRE-ATLANTIQUE



AGIR pour la
BIODIVERSITÉ
ANJOU



AGIR pour la
BIODIVERSITÉ
PAYS DE LA LOIRE



Table des matières

| | |
|--|----|
| ABSTRACT | 3 |
| RESUME | 3 |
| 1.INTRODUCTION | 4 |
| 2. MATERIEL ET METHODES | 8 |
| 2.1 Le site d'étude | 8 |
| 2.2. Les données | 9 |
| 2.2.1. Temporalité | 9 |
| 2.2.2. Les données sur la faune vivante | 9 |
| 2.2.3. Les données de mortalité de la faune par collision routière..... | 10 |
| 2.3. Prédicteurs (variables explicatives) | 10 |
| 2.3.1 Effort d'échantillonnage | 10 |
| 2.3.2. Danger (variables associées à la route) | 11 |
| 2.3.3. Covariables de paysage | 13 |
| 2.4. Le modèle mono espèce..... | 13 |
| 2.5. Résolution..... | 16 |
| 2.6. Validation..... | 17 |
| 3. RESULTATS..... | 18 |
| 3.1. Description des observations de chevreuils vivants et de cadavres | 18 |
| 3.2. Ajustement du modèle | 19 |
| 3.3. Estimation des paramètres..... | 21 |
| 3.4. Qualité prédictive | 22 |
| 4. DISCUSSION | 23 |

ABSTRACT

Roads are key vectors of the social and economic dynamism of a territory, but also represent a threat for wildlife, resulting in increased mortality due to vehicle collisions. Wildlife roadkills may cause serious safety issues for users and inflate mortality rates in animal populations. For both reasons, predicting the spatial distribution of roadkills along road networks has become a major concern for the ecological component of infrastructure developments in spatial territory planning. a predictive model of roadkills patterns can be envisioned as a risk analysis in which exposure is modelled through the predictors of a species' geographic distribution and effect through predictors of observed casualties along a road network. We evaluated the predictive performance of this framework by modelling jointly the spatial point patterns of living roe deer presences and roe deer roadkills in a INLA framework. We fitted our models with opportunistic geolocated records of living animals gathered by local naturalists and semi-protocolled records of roadkills performed daily by road patrols in Brittany, France. The fit and predictive quality of the model are unsatisfactory, but the plausibility of the predicted biological parameters leads us to consider further work to remove methodological barriers to propose a predictive model of vehicle-animal collisions with regional scope.

RESUME

Les routes sont des infrastructures clés du dynamisme d'un territoire. Elles représentent cependant une menace pour la faune entraînant un surplus de mortalité lié aux collisions avec les véhicules. Ces collisions représentent également un enjeu de sécurité routière pour les automobilistes. Pour ces deux raisons, pouvoir prédire la distribution spatiale des collisions véhicule-animal est un enjeu majeur pour le développement d'infrastructures d'aménagement du territoire. Un modèle prédictif des patrons spatiaux de mortalité routière peut être envisagé comme une analyse de risque dans laquelle l'exposition correspond à la distribution géographique d'une espèce et le danger à des caractéristiques relatives à la route. Nous avons évalué les performances prédictives dans ce cadre en modélisant conjointement le patron de présence spatiale des chevreuils vivants ainsi que la mortalité routière de ceux-ci via l'approche INLA. Pour ajuster notre modèle, nous prenons en compte à la fois des données opportunistes géolocalisées fournies par des associations naturalistes locales concernant les animaux vivants mais également des données semi-protocolées concernant la mortalité routière collectées par les patrouilles routières de Bretagne, en France. L'ajustement et la qualité prédictive du modèle sont insatisfaisants mais la crédibilité des paramètres biologiques prédits, pousse à envisager une poursuite pour lever les barrages méthodologiques afin de pouvoir proposer un modèle prédictif des collisions véhicule-animal à portée régionale.

1.INTRODUCTION

Les routes sont des infrastructures clef pour le dynamisme d'un territoire, mais constituent des ruptures fortes pour la connectivité écologique d'un paysage (Fahrig, 2003; Laurance & Balmford, 2013). Elles impactent les dynamiques de métapopulations animales à l'échelle des paysages en altérant la connectivité entre sous-populations, soit par une mortalité directe des propagules (effet direct), soit en induisant un effet comportemental répulsif (effet indirect), ces effets pouvant se combiner ou s'associer à des effets attractifs (Fahrig & Rytwinski, 2009; Jackson & Fahrig, 2011). Les routes induisent en particulier une surmortalité de certaines espèces de faune sauvage, principalement due aux collisions routières véhicule-animaux (Benítez-López *et al.*, 2010).

Face à cet enjeu de conservation, les politiques publiques cherchent à mitiger les effets des routes sur la faune afin maintenir la connectivité du territoire, par exemple à travers les trames vertes et bleues (Bergès *et al.*, 2010; Trame verte et bleue). Les conflits véhicules-faune intéressent de plus les politiques de sécurité routière lorsqu'il s'agit de grandes espèces, en particulier les ongulés (Clevenger *et al.*, 2015; Linnell *et al.*, 2020). Face à ce double enjeu, il apparait indispensable de disposer d'une cartographie du risque de conflits véhicules-faune, différenciée entre espèces à enjeux, et permettant de prioriser les actions d'atténuation à l'échelle d'un territoire et positionner des infrastructures d'évitement des collisions ou de facilitation des traversées de la faune (van der Grift *et al.*, 2013; Polak *et al.*, 2014; Grilo *et al.*, 2021).

La cartographie du risque de mortalité de la faune sauvage par collision routière est habituellement réalisée à partir de statistiques spatiales descriptives comme le K de Ripley ou l'estimation par noyau (KDE+) (Bíl *et al.*, 2013; Billon, 2018). La détermination des zones les plus mortifères consiste alors à identifier les zones comportant une densité d'observation de cadavres significativement plus élevée que celle attendue sous l'hypothèse d'une distribution aléatoire des événements de collision véhicule-faune (Favilli *et al.*, 2018). Bien que ces méthodes descriptives soient efficaces pour situer les points noirs de mortalité, elles ne permettent ni d'expliquer les patrons spatiaux des collisions véhicule-faune par des variables choisies sous hypothèses biologiques, ni de les prédire, contrairement aux approches par statistique inférentielle (Santos *et al.*, 2013; Chyn *et al.*, 2021). Les inférences par modélisation statistique sur la mortalité de la faune par collision routière peuvent être envisagées sous trois dimensions : spatiale, temporelle et spécifique (Clevenger *et al.*, 2003; González-Suárez *et al.*, 2018; Bíl *et al.*, 2019; Ignatavičius *et al.*, 2020; Medrano-Vizcaíno *et al.*). Les politiques publiques ayant prioritairement besoin d'une vision cartographique du risque moyen sur le temps long et à échelon territorial, la modélisation spatiale s'avère prioritaire sur les deux autres axes. Cet

exercice repose habituellement sur la détermination des prédicteurs de la mortalité liés à la route elle-même (nature, trafic, infrastructures) et à l'environnement (habitat environnant, structure du paysage) à partir de régressions logistiques et de modèles linéaires généralisés (Malo *et al.*, 2004).

A échelle régionale (dans notre cas, l'échelon territorial des politiques publiques, le plus souvent un découpage administratif quelconque), le processus de mortalité de la faune sauvage par collision peut être envisagé sous la forme d'une analyse de risque. Un risque est défini comme l'intersection entre un danger et une exposition (Papazoglou *et al.*, 2017) (Fig. 1). Dans le cas des collisions véhicule-faune, l'exposition correspond à l'occurrence d'un animal dans une zone traversée par une route, que l'on peut représenter comme l'intersection entre les routes et la distribution d'une espèce (Fig.1.a et Fig.1.c). Au niveau populationnel, l'exposition est un processus densité-dépendant, donc modulé par les variations locales d'occurrence et d'abondance des individus. Ce patron spatial correspond aux modèles statistiques de distributions d'espèces classiquement utilisés en écologie (« species distribution models » ou « habitat suitability models », SDM, (Guisan *et al.*, 2017), qui reposent sur une représentation synthétique de la niche écologique réalisée d'une espèce par des prédicteurs liés au climat, la topographie et l'habitat. L'exposition est de plus modulée par des caractéristiques propres à chaque espèce (traits écologiques), qui affectent à la fois sa relation à l'environnement (dans notre cas, l'habitat ou le paysage), sa mobilité et son comportement (Duffett *et al.*, 2020). Dans cette approche par SDM, l'exposition peut typiquement se représenter comme une surface de probabilité d'occurrence (comme dans les SDM basés sur des régressions logistiques, (MacKenzie *et al.*, 2002; Guillera-Aroita *et al.*, 2011)) ou d'intensité d'un patron de point (approche par modélisation de patrons de points, (Warton & Shepherd, 2010; Renner & Warton, 2013)). Le danger se comprend comme les propriétés de la route qui affectent la probabilité de collision à exposition donnée (Fig.1.b. et Fig.1.c). Les prédicteurs du danger correspondent à des caractéristiques intrinsèques à la route et à son environnement proche en un point de coordonnées donné, incluant des caractéristiques physiques (topologie, largeur, présence d'obstacles) et des données liées au trafic (vitesse, type de véhicule, nombre de véhicule), l'ensemble étant, comme l'exposition, modulé par les traits écologiques de l'espèce concernée (facteurs d'attraction ou de répulsion comme des modes de chasse ou de déplacement) (González-Suárez *et al.*, 2018). La modélisation statistique du danger repose principalement sur les variations du nombre de cadavres ou de leur probabilité d'occurrence en tout point d'une route, associées à des prédicteurs synthétiques des caractéristiques de la route.

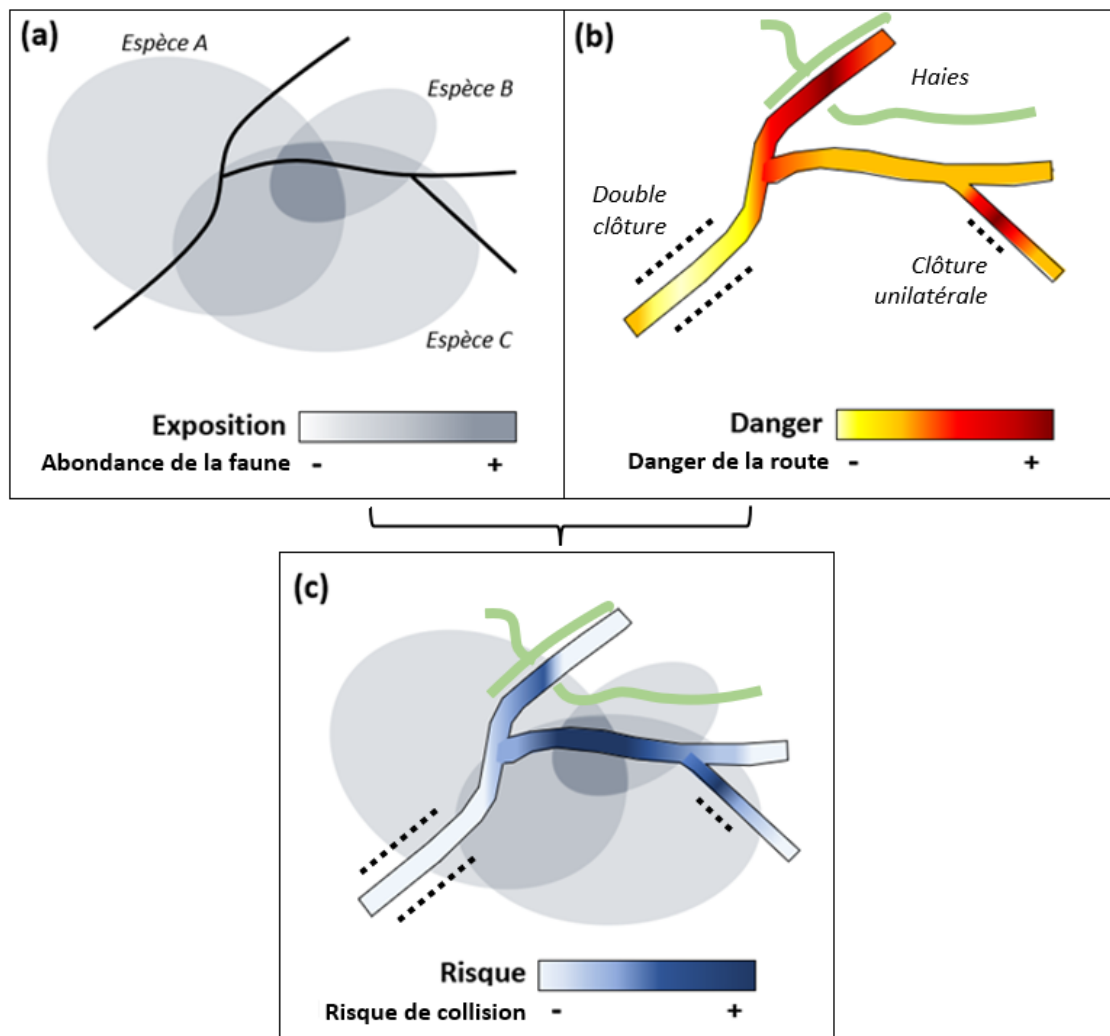


Figure 1 : Schéma de principe d'une analyse de risque couplant l'exposition (a), placement d'un animal de manière qui les expose aux effets de certains équipement, exprimée comme le recoupement de la distribution d'une espèce avec un réseau routier, et le danger (b), propriété intrinsèque d'un équipement susceptible d'avoir un effet nuisible, inféré à partir de descripteurs de la route. La résultante est une carte de prédiction spatiale du risque (c)

Afin de modéliser le risque de mortalité dans un objectif de cartographie prédictive, il est donc nécessaire de disposer de données indépendantes sur l'exposition des espèces et le danger lié à la route. A une échelle pertinente pour les politiques publiques, typiquement celle d'une région administrative, l'acquisition de données de distributions d'espèces est réalisable par des atlas en occurrence ou abondance semi-quantitative (Nos atlas régionaux - Bretagne Vivante), cependant limités dans leur capacité à représenter des gradients d'abondance à grain spatial fin. La modélisation du danger nécessite en revanche l'acquisition de données de mortalité (localisation de cadavres) inaccessible, à l'étendue spatiale d'un territoire, à des méthodes d'échantillonnage protocolés – à la fois lourdes à mettre en place et dangereuses pour les opérateurs. Ces limites imposent donc d'utiliser des données non-protocolées ou semi-protocolées, comme celles récoltées par les observatoires participatifs de biodiversité dont la plupart disposent désormais de modules de relevé géolocalisés tant pour les observations

d'individus vivants que morts (Billon *et al.*, 2015). Bien que ces données soient attractives par leur abondance et leur étendue spatiale, elles sont récoltées par des observateurs volontaires qui ne récoltent pas (ou rarement) de données d'absence. La distribution des observations est donc dépendante des variations spatiales de l'effort d'échantillonnage, qui est hétérogène et fonction des caractéristiques des observateurs (connaissance du terrain, préférences personnelles, lieu de vie, fréquences de passage sur le terrain) comme de l'accessibilité du terrain (altitude, accès publics, obstacles au passage, danger pour la sécurité des personnes) (Kelling *et al.*, 2015; Johnston *et al.*, 2021). Pour ce qui concerne les cadavres, il existe en France des données semi-protocollées, récoltées par les Directions Interdépartementales des Routes (DIR), qui pallient certains des biais liés à l'hétérogénéité des données opportunistes, tout en évitant d'inciter les observateurs à des comportements à risques. Les données des carcasses sont identifiées lors du passage de patrouilleurs des routes à environ 90km/h. Malgré les défauts inhérents à ces données (en particulier le fait qu'elles ne sont ni validées, ni récoltées par des naturalistes), elles permettent, pour certaines espèces facilement reconnaissables, de limiter l'impact de l'effort d'échantillonnage sur la composante « danger » d'un modèle de risque.

Dans cette étude, nous construisons le premier modèle prédictif spatialisé cartographiant le risque de mortalité de la faune sur la route à l'échelle d'un territoire, la région Bretagne. Ce travail s'inscrit dans une relation entre l'Etat (ministère en charge de l'écologie), des experts de l'aménagement du territoire (Cerema) et des écologues modélisateurs, et constitue à notre connaissance la première tentative de modélisation spatiale du risque de mortalité à une telle étendue spatiale. A cette fin, nous représentons explicitement la relation entre exposition et danger dans un cadre spatial par un modèle de distributions jointes (Escamilla Molgora *et al.*, 2022), en tenant compte des variations de l'effort d'échantillonnage. Notre travail est structuré par trois questions : (i) Quelle est la qualité descriptive d'un modèle spatial de distributions jointes pour les observations de mortalité par collision routière ? (ii) Quelle est la capacité prédictive de ce modèle dans l'espace et le temps ? (iii) Les paramètres décrivant l'exposition et le danger à partir de variables paysagères sont-ils biologiquement pertinents ? À étendue régionale, nous nous attendons à ce que le principal explicateur du risque de mortalité par collision routière de la faune soit l'exposition des espèces au danger routier (Zimmermann Teixeira *et al.*, 2017; Jasińska *et al.*, 2019). Si cette hypothèse se vérifie, le risque routier devrait être plus élevé dans les lieux comportant des habitats favorables aux espèces. Pour une même exposition, nous nous attendons à ce que le risque routier soit plus élevé pour les routes proches de la végétation, à fort trafic et à grande vitesse (Malo *et al.*, 2004; Girardet *et al.*, 2015).

2. MATERIEL ET METHODES

2.1 Le site d'étude

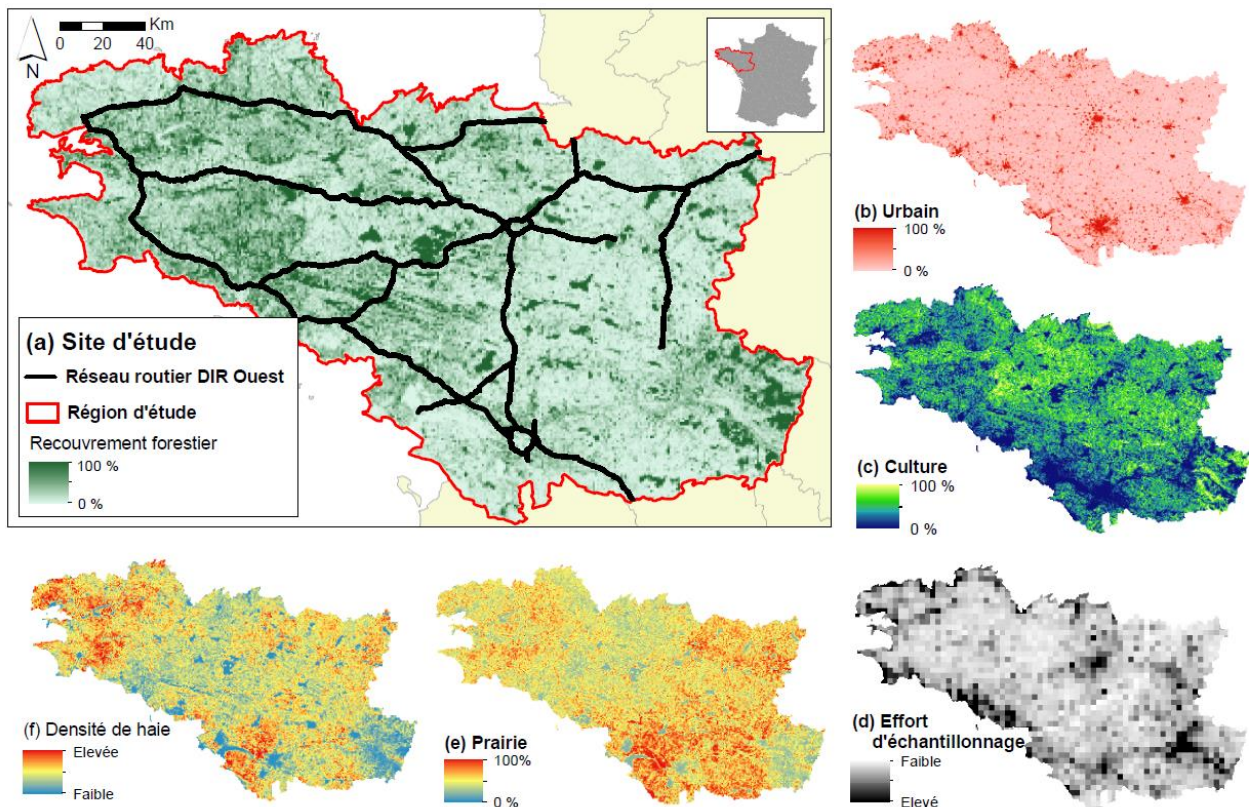


Figure 2 : Site d'étude et son paysage : (a) Région d'acquisition des données, réseau routier de la DIR Ouest et cartographie du recouvrement forestier par km^2 (b) Recouvrement des zones urbaines par km^2 (c) Recouvrement des cultures par km^2 (d) Cartographie de l'effort d'échantillonnage par maille $5 \times 5 \text{ km}^2$ (e) Recouvrement des prairies par km^2 (f) Densité de haie par km^2

La région étudiée recouvre le territoire de la Bretagne et trois départements des Pays de la Loire. Ce territoire est majoritairement constitué de paysages agricoles dont certaines portions ont conservé un bocage dense malgré le remembrement initié dans les années 1960. Le climat y est océanique et le relief modéré, les sommets les plus élevés avoisinant les 400 mètres. Ce territoire a été sélectionné pour le dynamisme des associations naturalistes qui permet une couverture spatiale, temporelle et taxonomique élevées par des observations opportunistes réalisées par des naturalistes amateurs (nous n'exploitons ici que les données d'animaux vivants pour une sélection d'espèces, voir 2.2.2.).

Ce territoire a également été choisi car la Direction Interdépartementale des Routes (DIR) Ouest a mis en place depuis 2014 un protocole visant à obtenir des informations sur les collisions routières de la faune qui nous permettent de disposer d'un jeu de données sur la mortalité entièrement indépendant des données de faune vivante (voir 2.2.3.). La DIR Ouest fait partie des DIR les plus assidues sur la récolte des collisions routières, stimulée par de nombreux partenariats et séances de formations à l'identification d'espèces avec la contribution du

CEREMA Ouest et des associations naturalistes locales. Le réseau couvert par la DIR Ouest représente 1590 km de route dont 1250 km de 2x2 voies.

La durée de la période étudiée débute en 2014 la première année de mise en place du protocole de comptage des carcasses par la DIR Ouest et s'étend jusqu'en 2021 de manière ininterrompue.

2.2. Les données

2.2.1. Temporalité

La durée de la période étudiée débute en 2014, la première année de mise en place du protocole de relevé des cadavres par la DIR Ouest, et s'étend jusqu'en 2021 de manière ininterrompue. Les données de faune vivante sont également disponibles sur cette période sans interruption. Dans notre modèle (voir 2.4.), nous n'incluons pas de dimension temporelle afin que la cartographie prédictive soit élaborée à partir de l'ensemble des données disponible comme un patron de « long terme ». La mortalité routière connaît une dynamique interannuelle et saisonnière (Raymond *et al.*, 2021), qui, outre qu'elle induit une complexité supplémentaire trop lourde à un stade précoce du développement du type de modèle que nous exploitons, ne nous intéresse pas particulièrement dans le cadre de cette étude précise.

2.2.2. Les données sur la faune vivante

Les données concernant la faune vivante proviennent de la base de données « Faune Bretagne », maintenue par Faune Bretagne, le groupe mammalogique Breton et la LPO Anjou, et qui participe au programme « Visionature » à travers le client de base de données et l'interface développés par la société BioloVision (Accueil - data.biolovision.net; Visionature). Ces données consistent en des observations géoréférencées, saisies sur le terrain par les observateurs via l'application NaturaList. Elles sont opportunistes, c'est-à-dire qu'elles ne suivent (sauf dispositif spécifique non pris en compte dans cette étude) aucun protocole : l'observateur choisit donc de saisir ou non une donnée en fonction de ses préférences personnelles. Ces données consistent exclusivement en des points de présence géolocalisés associés à une espèce, une date et éventuellement un nombre d'individus. Bien que très déterminées par la distribution des observateurs et l'effort d'échantillonnage (voir section 2.3.1.), les données opportunistes restent la source d'information sur les distributions d'espèces la plus fournie et accessible pour l'étendue spatiale visée. Des atlas de présence sont également disponibles (Nos atlas régionaux - Bretagne Vivante) et ont l'avantage d'un protocologage au moins partiels, mais ils n'offrent qu'une image diachronique des occurrences (éventuellement exprimées en classes d'abondances) d'espèces, discrétisées en cellules relativement grossières. Les données opportunistes donnent accès à des résolutions temporelles plus fines (en temps continu, jusqu'au jour voire l'heure) et, si l'effort d'échantillonnage est correctement pris en compte, à des gradients spatiaux d'abondance ou

d'intensité de présence relatives. Nous avons récupéré pour les années 2014 à 2021 : l'identification à l'espèce de l'animal observé, les coordonnées de sa position, la date d'observation, l'identifiant de l'observateur, la précision de l'observation (les données non géolocalisées ont été exclues), l'état de l'animal (animaux morts exclus).

2.2.3. Les données de mortalité de la faune par collision routière

Les données de mortalité par collision sont récoltées quotidiennement par les agents DIR Ouest en sus de leurs tâches principales de surveillance du réseau routier. Au cours de leur patrouille, les agents identifient les carcasses aperçues sur la partie imperméabilisée de la route, et en notent la position comme une distance à un point repère (précision de 100 à 200m due à la vitesse des patrouilles). Pour des raisons de sécurité, les patrouilles sont réalisées quotidiennement à une vitesse de 60 à 100km/h, selon le type de route

Les données obtenues par ce protocole DIR comportent l'année et le mois d'observation, les coordonnées géographiques du cadavre ainsi que son nom vernaculaire. L'identification des carcasses est accessible via 3 champs qui peuvent indiquer un groupe taxonomique, un nom d'espèce ou un groupe arbitraire (rapaces nocturnes, autre petite faune, etc.). Il arrive que l'information portée par ces 3 champs soit contradictoire (par exemple, un point de donnée peut être renseigné comme « Mustelidés » avec comme nom d'espèce « Buse »). Ces incohérences résultent du fait que les agents DIR ne sont pas spécifiquement formés à l'identification et n'ont, pour la plupart, aucune compétence ni intérêt naturaliste. En cas d'incohérence manifeste, nous avons supprimé les points de données lorsque deux champs différents portaient une précision à l'espèce différente et lorsque le groupe d'espèce et l'espèce observée étaient soit « groupe de mammifère – espèce d'avifaune », soit « groupe d'avifaune – espèce de mammifère ». Nous avons conservé les données, avec les corrections nécessaires, dans les cas où l'identification de l'espèce s'avère plus transparente malgré des erreurs de champs.

2.3. Prédicteurs (variables explicatives)

2.3.1 Effort d'échantillonnage

Les données sur la faune vivante doivent être analysées au regard des processus d'observation des données opportunistes. Une observation opportuniste dépend de la combinaison entre des caractéristiques de site (accessibilité, attractivité) et des caractéristiques de l'observateur (heure d'observation, durée d'observation, capacité d'identification des espèces) (Kelling *et al.*, 2015; Johnston *et al.*, 2021). Après avis des gestionnaires de la base de données Faune Bretagne, nous avons représenté l'effort d'échantillonnage par quatre variables sur une maille de 5 x 5 km pour l'ensemble du territoire :

- le nombre total de point d'observation, toutes espèces confondues (reflet de la pression d'observation sur une maille),
- le nombre de dates comportant des observations (reflet de la fréquence de passage sur une maille),
- le nombre d'espèces observées sur la maille (reflet de la diversité des naturalistes sur la maille et de la richesse spécifique de la maille)
- le nombre d'observateurs différents (reflet de l'hétérogénéité de la récolte de données)

Ces quatre variables sont fortement corrélées, avec un r^2 allant de 0.47 (nombre d'espèces observées - nombre d'observateurs) à 0.86 (nombre de date uniques - nombre de données). Pour éviter la colinéarité, nous avons réalisé une Analyse en Composantes Principales (ACP) centrée réduite afin de résumer la pression d'observation en deux axes synthétiques (respectivement 74.6% de la variance totale pour le premier axe PC1 et 14,6% pour le deuxième axe PC2). PC1 est négativement corrélé aux 4 variables (de -0.72 à -0.93). PC2 est positivement corrélé au nombre d'espèce observées (0.69) et négativement corrélé au 3 autres variables (de -0.14 à -0.27). Nous n'avons finalement conservé que le premier axe. Pour des raisons de facilité d'interprétation, nous avons inversé les valeurs de PC1 dans la covariable *sampling* (eq. 1). Plus les valeurs de *sampling* sont élevées, plus l'effort d'échantillonnage est élevé.

$$sampling = -1 \times Axe1 \quad (eq. 1)$$

Cette covariable comprend un nombre réduit de valeurs extrêmes. Afin d'équilibrer le poids des valeurs de la covariable *sampling*, nous avons réalisé une transformation racine carrée une fois la valeur absolue de son minimum ajoutée pour que la valeur minimale de *sampling* ne soit pas inférieure à zéro (afin de pouvoir appliquer une transformation racine) (eq. 2).

$$SAMPLING = \sqrt{sampling_s + |\min(sampling_s)|} \quad (eq. 2)$$

2.3.2. Danger (variables associées à la route)

Les données sur le trafic routier sont fournies par la DIR Ouest et contiennent le trafic moyen journalier total sur l'année 2020 en véhicules jour, établi par le logiciel IRIS (logiciel internet de traitements des données de comptages). Elles sont interpolées sur l'ensemble du réseau DIR à partir de données provenant de stations de comptage (boucles inductives) implantées sur le réseau. Nous avons moyenné ces valeurs interpolées (en incluant véhicules légers et lourds) dans un raster de maille 1x1km.

Les données sur la vitesse proviennent de la BD TOPO (accès libre source : IGN mars 2021, <https://geoservices.ign.fr/telechargement>, fichier de forme : TRONCON_DE_ROUTE.shp)

et fournissent la vitesse moyenne des véhicules automobile légers par tronçon de route. Cette vitesse est estimée à partir de la nature des route (autoroute, bretelle, route nationale...) et de leur importance pour le trafic routier (liaison entre grande métropoles, départements, grosses agglomérations, communes, etc.). J'ai moyenné la vitesse dans un raster de maille 1x1km. Les vitesses moyennes du réseau DIR Ouest se distribuent selon 2 grandes classes ; les vitesses inférieures ou égales à 90km/h, et les vitesses supérieures à 90km/h. Nous avons donc décidé de considérer la vitesse comme une variable qualitative à 2 classes : vitesse \leq 90km/h nommée *vitesse90* et vitesse $>$ 90km/h nommée *vitesse 100* .

Nous avons considéré que l'environnement local (sur quelques mètres de chaque côté de la route), caractérisé par l'habitat local et la présence de corridors écologiques, contribue à son danger pour la faune. S'il pourrait aussi être considéré comme une strate locale de l'exposition, nous avons choisi de l'associer au danger afin de faciliter la construction du modèle. Nous avons retenu comme descripteurs de l'environnement local la distance à la végétation la plus proche et la distance à un cours d'eau. Ces variables représentent la proximité de la route à un potentiel corridor écologique et, pour la distance à la végétation, elle constitue un indicateur de la proximité à l'habitat local de certains oiseaux (Benítez-López *et al.*, 2010). Pour mesurer ces distances, nous avons récupéré les données sur l'hydrologie et la végétation à partir de la BD TOPO (source : IGN mars 2021, fichier de formes : HAIE.shp - ZONE_DE_VEGETATION.shp - TRONCON_HYDROGRAPHIQUE.shp).

Les données sur la végétation comprennent, par département, le linéaire des haies (formations linéaires comportant des arbres, arbustes ou arbrisseaux sur une largeur inférieure à 20 m) et les polygones correspondant aux espaces végétaux naturels ou non. Pour obtenir la donnée de distance à la végétation, nous avons calculé dans un tampon de 2km autour de la route, un raster 500x500m qui indique la distance minimale du centroïde d'un pixel avec le linéaire de haie et la végétation.

Les données sur les cours d'eau comprennent, par département, un linéaire correspondant à l'axe du lit d'une rivière, d'un ruisseau ou d'un canal (source). Les tronçons hydrographiques sont classés selon leur numéro d'ordre qui en exprime le degré de ramification. Nous avons conservé les cours d'eau majeurs (numéro d'ordre de 4 ou plus). Pour obtenir la donnée distance à ces cours d'eau, nous avons calculé, dans un tampon de 2km autour de la route, un raster 50x50m qui indique la distance minimale du centroïde d'un pixel au cours d'eau le plus proche.

2.3.3. Covariables de paysage

Toutes les données de paysage ont été calculées à une résolution de 1x1km sur l'étendue de la région d'étude (Fig. 2). J'ai calculé la densité de haie par pixel à partir des fichiers de forme linéaire de haie provenant de la BD TOPO (source : IGN mars 2021, fichier de forme : HAIE.shp). À partir de la carte d'occupation du sol de la France métropolitaine (raster 10x10m), nous avons calculé le recouvrement de prairie (13), culture (5,6,7,8,9,10,11,12), forêt (16,17), et zone urbaine (1,2,3) (Source : OSO 2020, lien : <https://www.theia-land.fr/>).

2.4. Le modèle mono espèce

Nous disposons des données de faune vivante et de collision pour les 27 espèces de mammifère et d'avifaune les plus observées sous forme de carcasse par les agents de la DIR Ouest. Nous avons retenu le chevreuil comme espèce test, car il présente diverses caractéristiques qui facilitent les premiers développements du modèle (espèce présente partout sur le territoire mais avec des modulations d'abondance, facilement reconnaissable, facilement détectable sur les routes par les agents des DIR). Un modèle multi-espèces est cependant en cours de développement (Cf . Encart 1).

Nous avons analysé le risque de mortalité par collision routière des chevreuils en modélisant conjointement les observations d'individus vivants (Y^{vivant}) et les observations des carcasses sur la chaussée ($Y^{collision}$) sans effet de guildes. Nous avons modélisé ces deux jeux de données comme étant tirés chacun dans une loi de poisson inhomogène qui admet que l'intensité λ varie spatialement (eq. 2 et 3) (Renner *et al.*, 2015).

$$Y_s^{vivant} \sim P(E_s * Pobs_s) \quad (eq. 2)$$

$$Y_s^{collision} \sim P(E_s * D_s) \quad (eq. 3)$$

Où : s est un point défini par un jeu de coordonnées spatiales xy (projection Lambert 93 dans nos données et pour nos prédictions cartographiques).

Les intensités des processus de poisson inhomogènes des observations des vivants et observations des carcasses peuvent être décomposée en trois variables latentes (variable estimée à partir de covariables observées) ; E_s , $Pobs_s$ et D_s .

- La variable latente E_s correspond à l'exposition des chevreuils et intervient dans la modélisation des individus vivants comme dans celle des carcasses (eq. 2 et 3). C'est une modélisation de la distribution de l'espèce à partir de la niche écologique réalisée, exprimée par des prédicteurs paysagers (analogue aux *habitat suitability models* élaborés sous Maxent et bien connus des écologues, (Renner & Warton, 2013)). Le logarithme de

E_s est modélisé à partir d'une combinaison de relations linéaires avec les covariables de paysage (voir section 2.3.3 du Matériel et Méthodes et eq. 4). L'approximation E_s correspond à l'intensité de présence de chevreuils qui peut s'interpréter comme une abondance relative de chevreuils.

$$\ln(E_s) = \alpha + \beta^{haie} . HAIE_s + \beta^{culture} . CULTURE_s + \beta^{foret} . FORET_s + \beta^{prairie} . PRAIRIE_s + \beta^{urbain} . URBAIN_s \quad (eq. 4)$$

- La variable latente $Pobs$ est la probabilité d'observer un chevreuil vivant (eq. 2), modélisée comme une régression logistique de la covariable de synthèse d'échantillonnage $SAMPLING_s$ pondérée par un paramètre $\beta^{sampling}$ (voir section 2.3.1 du Matériel et Méthodes et eq. 5). Nous avons forcé le paramètre $\beta^{sampling}$ à être supérieur ou égal à 0 puisque des valeurs élevées de la covariable $SAMPLING_s$ induisent une probabilité élevée d'observer un animal ; ce paramètre sert donc seulement à déterminer la forme de $Pobs_s$:

$$Pobs_s = \text{logit}(\beta^{sampling} . SAMPLING_s) \quad (eq. 5)$$

- La variable latente D_s correspond à la variabilité de l'occurrence des carcasses de chevreuil observées (danger, eq. 3). Le logarithme de D_s est modélisé par une combinaison de relation linéaires entre les covariables de route quantitatives et d'un intercept $\alpha^{VITESSE_s}$ qui dépend de la classe de vitesse de la route à la localisation s (voir section 2.3.2. du Matériel et Méthodes et eq.6).

$$\ln(D_s) = \alpha^{VITESSE_s} + \beta^{trafic} . TRAFIC_s + \beta^{dist_veg} . DIST_VEG + \beta^{dist_hydro} . DIST_HYDRO \quad (eq. 6)$$

1 : Travail prévisionnel sur un modèle multi espèce

Guildes écologiques

Les traits écologiques regroupant des traits comportementaux et biologiques des espèces sont déterminant dans l'intensité de leur exposition au danger routier. Cette influence provient de traits comportementaux comme de traits morphologiques. Nous avons déterminé des groupes d'espèces susceptibles de posséder une similarité par rapport au risque routier.

Nous avons discriminé des guildes pertinentes au regard du risque routier à partir de divers traits écologiques. Les premiers sont le poids et la taille moyenne des espèces, qui impactent la détection par les usagers de la route, et qui sont reliés au déplacement des espèces, généralement plus une espèce est grande plus son aire de vie sera grande (Jetz *et al.*, 2004). Nous avons ensuite intégré le régime alimentaire, et le type de comportement de recherche de nourriture qui modulent la fréquence et l'aire des déplacement des espèces ainsi que leur attraction à la route (carnivores - charognards). Le déplacement des espèces est également caractérisé par leur grégarisme, leur vitesse pour les mammifères et leur type de vol (battu, plané) pour l'avifaune. La nocturnité des espèces impacte quant à elle majoritairement l'évitement par les conducteurs, qui ne perçoivent pas l'arrivée d'un animal assez rapidement.

À partir des traits écologiques cités, nous avons construit une matrice de distance entre espèce à partir de l'indice de Gower (Kaufman & Rousseeuw, 2009), séparément pour les mammifères et pour l'avifaune. L'indice de Gower permet d'intégrer des variables qualitatives et quantitatives dans un même calcul de distance. Nous avons ensuite réalisé une classification hiérarchique ascendante par saut maximal (critère d'agrégation = complete). À partir de cette analyse nous avons discriminé 6 groupes.

L'intégration d'un effet guildes au modèle monospécifique est envisagé comme suivant :

Les paramètres α et β sont dépendants de l'espèce i , ils sont décomposés en une moyenne de guildes et un effet aléatoire de l'espèce. La somme des effets de toutes les espèces d'une même guildes est nulle afin que les effets espèces soient identifiables. Par exemple pour la pente de l'effet forêt :

$$\beta_i^{foret} = B^{foret} + b_i^{foret}$$

$$\sum_{i=1}^{ni} b_i^{foret} = 0$$

Où B^{foret} est la moyenne de guildes et b_i^{foret} l'effet aléatoire de l'espèce i .

2.5. Résolution

L'estimation des paramètres α et β de ce modèle a été réalisée dans un cadre Bayésien qui admet chaque paramètre comme une variable aléatoire dont on cherche la distribution a posteriori. La distribution a posteriori des paramètres du modèle correspond à l'ajustement de leur loi a priori à partir des observations de chevreuils vivants et morts par collision routière. Les paramètres α et β du modèle suivent tous une loi normale non informative puisque nous n'avons intégré aucune information a priori concernant l'impact des covariables sur l'occurrence des individus vivants et des carcasses.

Dans le cadre bayésien, l'estimation de la loi jointe a posteriori d'un modèle est généralement calculée par une méthode numérique itérative MCMC (Monte Carlo Markov Chain). Cependant, si la méthode MCMC fournit une solution asymptotiquement exacte, elle est coûteuse en temps de calcul. Une solution moins coûteuse est l'utilisation de la méthode INLA (Integrated Nested Laplace Approximation (Rue *et al.*, 2009; Opitz, 2017)), une approximation analytique des lois a posteriori (Cf. Encart 2). Après des essais préliminaires en MCMC qui se sont avérés non concluants d'un point de vue pratique (temps de calcul excessif), nous avons choisi une résolution par la méthode INLA, pour se faire nous avons implémenté le modèle sous RStudio et utilisé les packages INLA et inlabru. Les paramètres ont été estimés à partir d'un maillage (triangulation de Delaunay) dont les points sont espacés au maximum de 5km, la résolution des prédictions est donc d'environ 25km². Les paramètres permettent de définir l'intensité en un

2 : INLA (Integrated Nested Laplace Approximation)

La méthode INLA est une inférence bayésienne approximative qui peut être appliquée aux modèles gaussiens latents.

Un modèle gaussien latent est un modèle dont la variable réponse y suit une loi exponentielle et sa moyenne μ est formée de composantes latentes gaussiennes. Cette condition est vérifiée si :

- la moyenne μ est liée par une fonction de lien à un prédicteur additif (addition des effets de multiples covariables)
- la distribution des paramètres associés aux effets des covariables est gaussienne (loi a priori gaussiennes).

Dans notre cas, les variables d'observations Y suivent une loi de Poisson, leur moyenne λ est composé de variables latentes qui sont des prédicteurs additifs de l'effet des covariables et les paramètres modulant l'effet des covariables ont des loi a priori Gaussienne.

Les loi a posteriori marginales des paramètres sont ensuite estimées à partir d'une approximation de Laplace (approxime une distribution sur la base d'une extension du théorème de Taylor).

L'estimation des loi a posteriori marginale se fait à partir d'une partition du champ des composantes latentes (partition définie par la triangulation de Delaunay), pour chaque point d'intersection de la partition. La prédiction continue de la moyenne μ sera alors issue de l'intégration des estimations ponctuelles.

point donné, cette intensité doit ensuite être interpolée pour obtenir une estimation du nombre de chevreuils par pixel (Intensité de chevreuils dans un pixel de 24 km² : multiplication de l'intensité du centroïde du pixel par l'aire d'un pixel – Intensité de cadavre dans un pixel de 24km² : multiplication du centroïde du pixel par la longueur de la portion de route traversant ce pixel).

2.6. Validation

L'ajustement du modèle aux données utilisées pour estimer ce paramètre a été évaluée au regard des indices suivants : PBIAS (Percentage of BIAS), RMSE (Root Mean Square Error). Le PBIAS correspond au pourcentage de biais moyen de la prédiction sur l'ensemble des observations, proche de zéro il indique une prédiction non biaisée, son signe indique si les prédictions ont une tendance à la sur- ou la sous-estimation (eq. 7). La précision de la prédiction sera évaluée par le RMSE, la racine de l'écart quadratique moyen entre les données simulées (intensité dans notre cas) et les données observées, proche de 0 il indique une précision fine des données simulées par rapport aux données observées (eq. 8). Il indique la variabilité de la qualité de la prédiction, afin de faciliter son interprétation, il a également été exprimé comme un pourcentage de l'observation moyenne (eq. 9).

$$PBIAS = 100 \times \frac{\sum(sim-obs)}{\sum obs} \quad (eq. 7)$$

$$RMSE = \sqrt{\frac{\sum((sim-obs)^2)}{nobs}} \quad (eq. 8)$$

$$RMSE.percent = \frac{RMSE}{\frac{\sum obs}{nobs}} \times 100 \quad (eq. 9)$$

La qualité prédictive temporelle et spatiale du modèle a été évaluée par validation par blocs et mesurée à partir des indices cités précédemment (Roberts *et al.*, 2017).

Pour la prédiction temporelle, nous avons estimé les paramètres du modèle de training à partir de données issues de 5 années consécutives (7500 points d'observation de chevreuil vivants et 1450 points de données de cadavre). Nous avons ensuite mesuré la qualité prédictive de ce modèle sur une année non adjacente aux années de training, sur 1500 points de chevreuil vivants et 290 points de cadavres (les intensités prédites par le modèle de training ont été divisées par 5). Nous avons réitéré cette opération 4 fois : [test : 2014 – training : 2016 à 2020], [test : 2015 – training : 2017 à 2021], [test : 2020 – training : 2014 à 2018], [test : 2021 – training : 2015 à 2019].

Pour la prédiction spatiale nous avons estimé les paramètres du modèle de training sur 3 blocs adjacents de 3600 km² chacun et comportant au minimum 20km de route (1750 points d'observations de chevreuils vivants et 375 points de cadavres). La qualité prédictive de ce

modèle a été mesurée sur un bloc non adjacent aux trois blocs de training, sur 350 points de chevreuils vivants et 75 points de cadavres (les intensités prédites par le modèle de training ont été divisées par 5). Nous avons réitéré cette opération 6 fois.

3. RESULTATS

3.1. Description des observations de chevreuils vivants et de cadavres

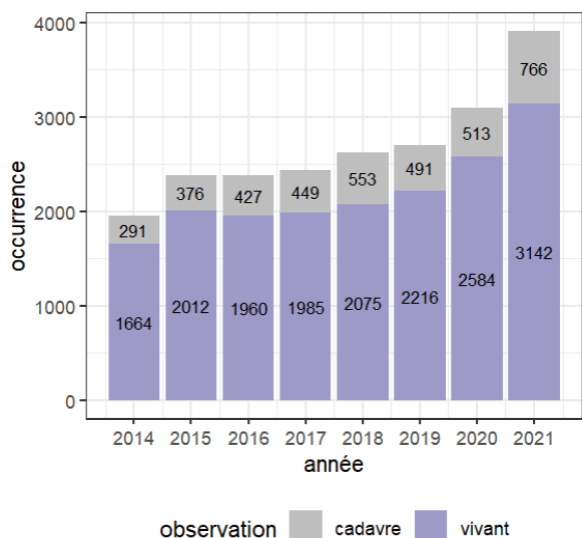


Figure 3: Variation annuelle de l'occurrence des chevreuils vivants sur la région d'étude et de cadavre sur le réseau de la DIR Ouest

Entre 2014 et 2021, 17107 chevreuils vivant ont été géoréférencés, soit en moyenne 2 138 chevreuils par an (Fig. 3). Pour les cadavres issus de collisions routière, en moyenne 482 cadavres ont été observés par année, ce qui représente 1.7 cadavre de chevreuil pour 10km par an (Fig. 3).

Les zones riches en observation de chevreuils vivant sont situées principalement dans les Pays de la Loire à proximité d'Angers, Laval et la route les reliant mais également au nord de Rennes et à Vanne (Fig. 4.a). La carte est réalisée à une résolution de 24km² et comporte quelques pixels qui regroupent une très grosse quantité de données (de 100 à 200 observations en 8 ans) (Fig. 4.a).

Le nombre cadavres de chevreuil collectés sur les routes varie entre 0 et 30 pour 24km². Les zones les plus mortifères se situent au Nord de Rennes, sur les routes partant de Lorient (Fig. 4.b).

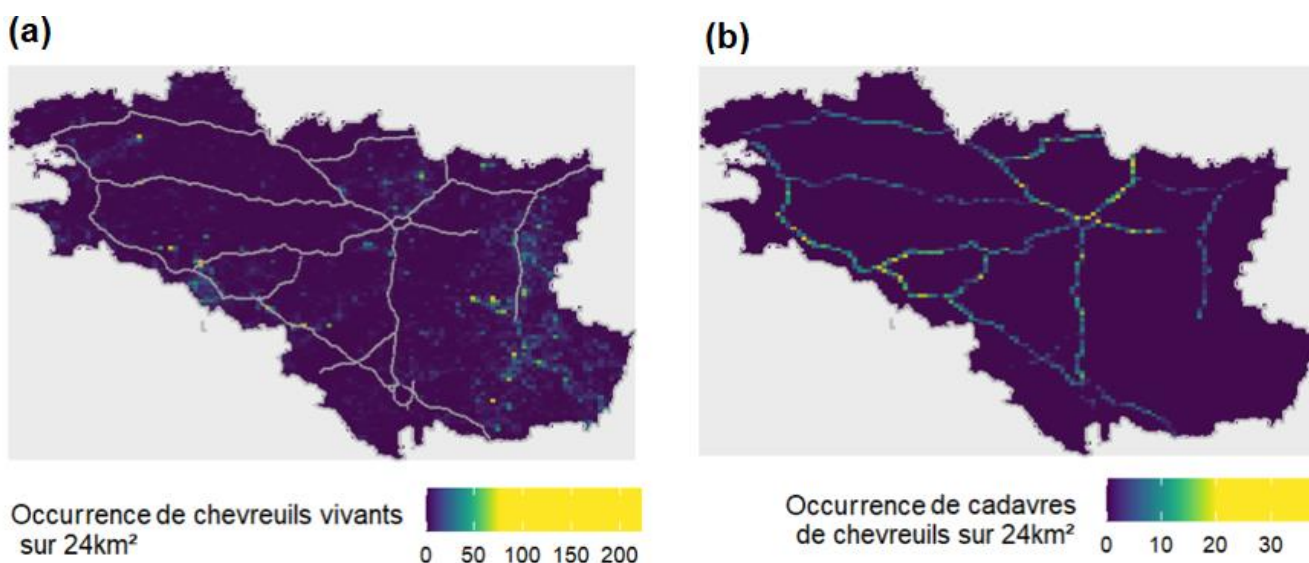
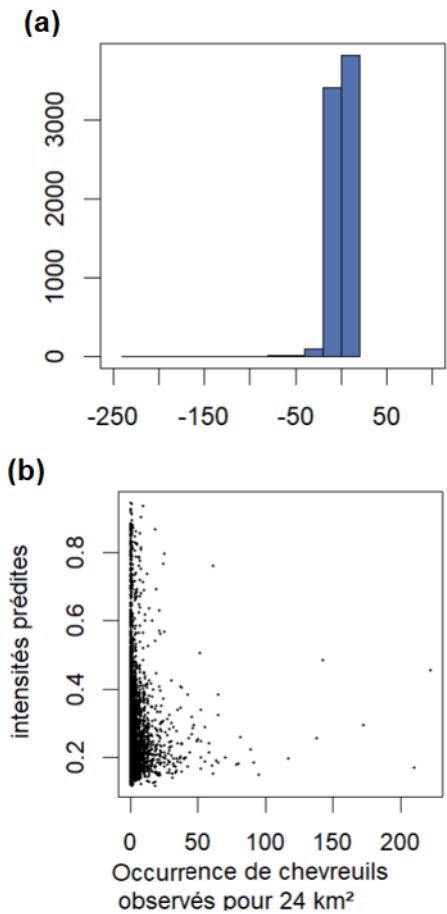


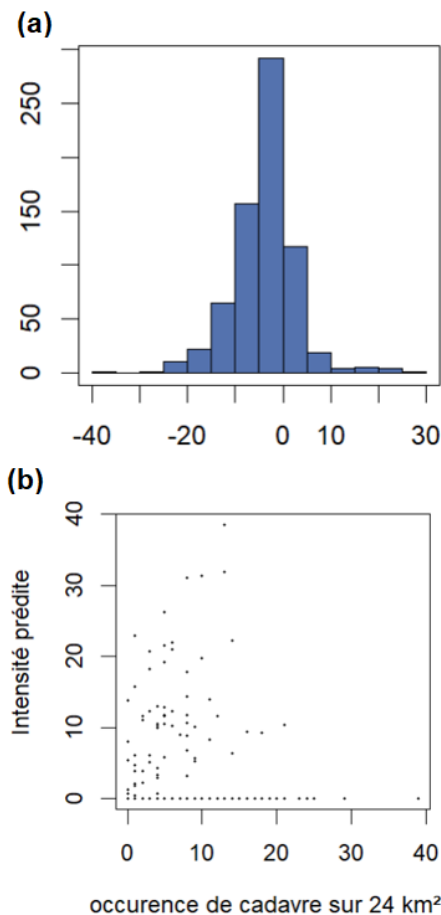
Figure 4 : Observations des chevreuils sur la période de 2014 à 2021, (a) vivants par par le programme Vlsionature, (b) à l'état de cadavre sur le réseau de la DIR Ouest par les patrouilleurs

3.2. Ajustement du modèle



L'intensité prédite la distribution observée de chevreuils vivants (Y^{vivant}) est sous-estimée de 88% en moyenne par rapport aux observations (PBIAS = -88.32). La RMSE associée à cette prédiction est de 7.81, l'erreur quadratique moyenne représente 336% de la valeur moyenne des observations, la prédiction est donc très imprécise. Les résidus ($donnee.predite - donnee.observee$) se distribuent principalement autour de zéro possèdent une grande queue décalée vers la gauche traduisant une tendance à la sous-estimation (Fig. 5.a). La gamme de variation des intensités de chevreuils prédites et des occurrences d'observation sur 24km² n'est pas comparable (de 0 à 1 pour les prédictions et de 0 à 200 pour les observations). Aucune tendance ni corrélation ne se dégage entre les intensités prédites et l'occurrence des observations de chevreuil vivant (Fig. 5.b).

Figure 5 : Ajustement du modèle pour la prédiction de la distribution observée des chevreuils, (a) distribution des résidus, (b) Relation entre prédiction et observations



L'intensité prédite des cadavres chevreuils géoréférencés ($Y^{collision}$) est sous-estimée de 80% en moyenne (PBIAS = -80.13) par rapport aux observations. Son RMSE est de 7.92, il représente 142% de la valeur moyenne des observations, la prédiction est imprécise. Les prédictions des cadavres possèdent la même gamme de variation que les observations de cadavres (entre 0 et 30 cadavres pour 24 km²) (Fig. 7c). Cependant les intensités de cadavre prédites sont plus clivées que les observations, on observe en effet une casi-absence de valeurs d'intensité intermédiaires (Fig. 7.c). Les résidus ont une distribution semblable équilibrée autour de la moyenne qui est légèrement inférieure à 0 (Fig. 6.a).

Figure 6 : Ajustement du modèle pour la prédiction des cadavres de chevreuils observés sur le réseau de la DIR Ouest, (a) distribution des résidus, (b) Relation entre prédiction et observations

En dehors des valeurs d'intensité prédites comme nulles qui correspondent à des occurrences des cadavre de chevreuils allant de 0 à 40 observations (ligne horizontale sur l'abscisse), une relation positive se dégage entre les intensités prédites et les observations de cadavres bien que la corrélation ne soit pas significative (Fig. 6.a).

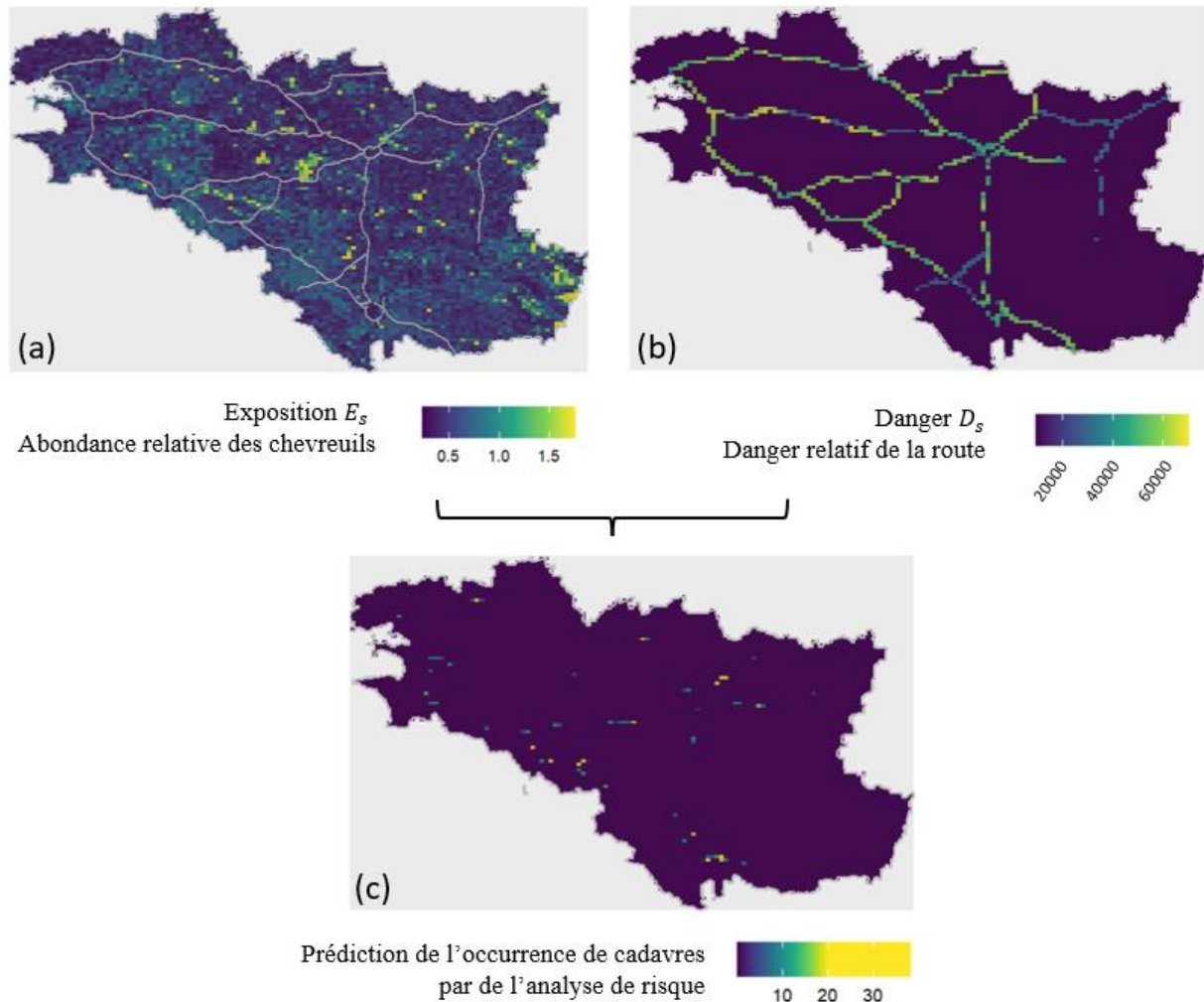


Figure 7 : Prédiction des collisions chevreuil-véhicule par l'analyse de risque entre 2014 et 2021 à résolution de 24km², (a) Prédiction de l'exposition E_s soit l'abondance relative de chevreuil déterminée par le modèle de niche, (b) Prédiction du danger D_s intrinsèque à la route, (c) Prédiction de l'occurrence de cadavres de chevreuils $Y_s^{collision}$

L'abondance relative du chevreuils E_s est répartie de manière relativement uniforme sur le territoire par rapport aux occurrences de chevreuils observées, avec une intensité très élevée au niveau des forêt (Fig. 7.a et Fig. 1.a). Le danger intrinsèque à la route D_s est plus élevé sur les portions de routes supérieures à 90km/h. Il est faible aux alentours des villes, soit sur les périphériques qui comportent un fort trafic (Fig. 7.b). La prédiction des cadavres issus de l'analyse de risque $Y_s^{collision}$ comporte majoritairement des prédictions élevées. Les points notables où une forte occurrence de collision chevreuil-véhicule est prédite sont la route entre Lorient et Vannes, l'autoroute un nord de Rennes et une partie du périphérique Sud de Nantes,

seule la partie du périphérique Sud de Nantes ne correspond pas à un espace identifié comme plus mortifère sur la cartographie des observations (Fig. 7.c et Fig.4.b).

3.3. Estimation des paramètres

- La relation entre la probabilité d'observation de chevreuil vivant et la covariable *SAMPLING* est assimilable à une relation linéaire positive dont l'intercept (pour les valeurs minimums de *SAMPLING*) est de 0.45 soit 45% de chance d'observer les chevreuils présents. La probabilité d'observation est donc relativement toujours élevée sur le territoire (Fig. 8).

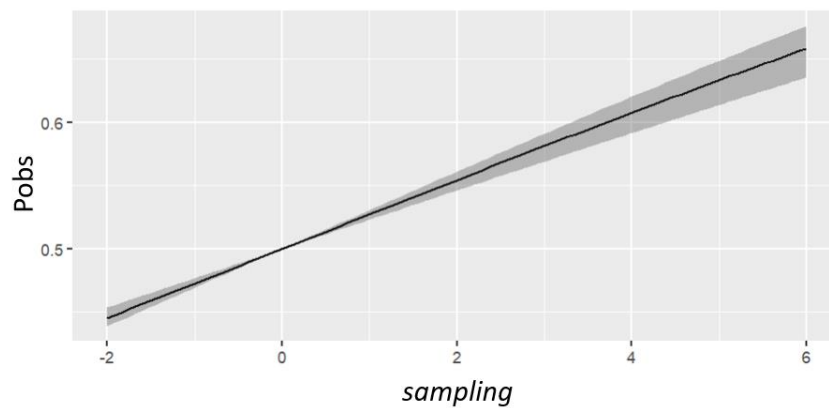


Figure 8 : Relation entre la probabilité d'observer un chevreuil vivant et la covariable d'échantillonnage *sampling*

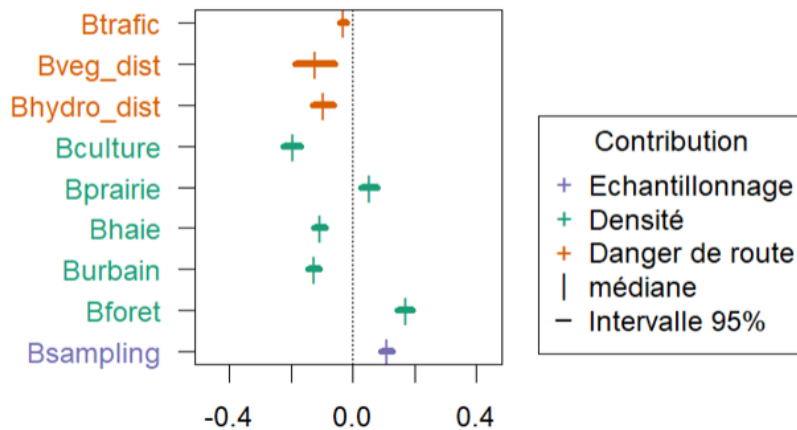


Figure 9 : Estimation des pentes des effet linéaires des covariables sur le danger de la route, la densité relative des chevreuils (exposition) et l'échantillonnage (les B correspondent à β)

- Pour les paramètres d'exposition, l'intercept α est très faible avec une médiane de -15.334 et un intervalle 95% : [-15.351 ; -15.318], impliquant que la densité de chevreuil soit faible pour un point. Les forts recouvrements culture, de zones urbaines et une densité de haie élevées influent négativement sur l'abondance relative de chevreuils (Fig. 9). Une

augmentation du recouvrement de prairie et de forêt implique en revanche une abondance relative de chevreuils plus élevée (Fig. 9).

- Les paramètres liés au danger intrinsèque à la route D_s , comportent un intercept quand la vitesse est inférieure ou égale à 90km/h, $\alpha^{vitesse90}$, avec une médiane de 10.181 et un intervalle de confiance de 95% de [10.076 ; 10.283] tandis que l'intercept des routes de vitesse supérieure à 90km/h, $\alpha^{vitesse100}$, a une médiane de 11.126 et un intervalle de confiance de 95% de [11.054 ; 11.197]. L'intercept des routes de vitesse supérieures est donc plus haut que celui des routes de vitesses inférieure, ce qui signifie que le danger routier est plus élevé sur les routes à grande vitesse (Fig.9). Le trafic a une relation négative avec le danger de la route. La distance à la végétation et la distance au cours d'eau le plus proche est reliée négativement au danger routier, ce qui signifie que les portions de routes à proximité de la végétation et des cours d'eau sont plus dangereuses (Fig. 9).

3.4. Qualité prédictive

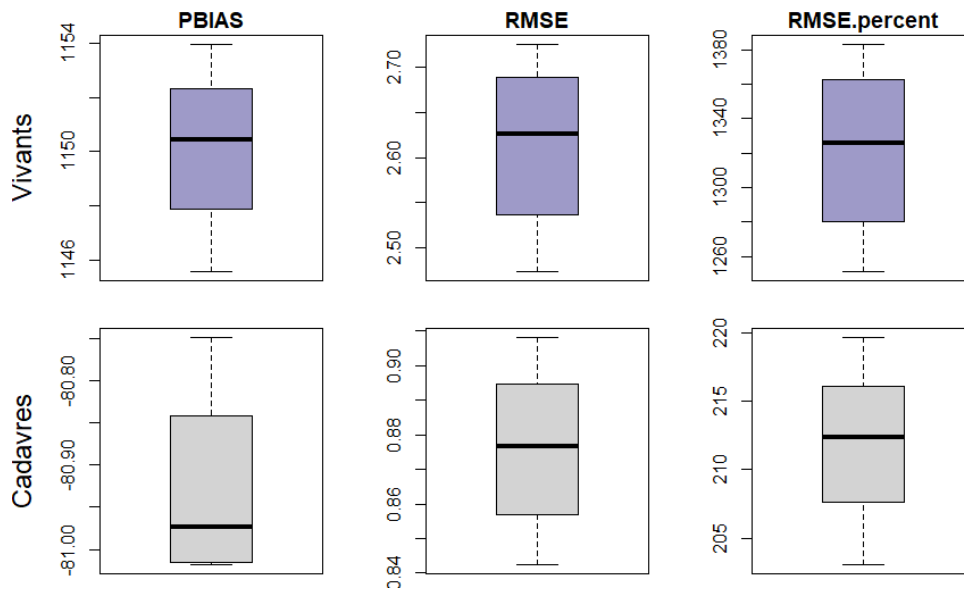


Figure 10 : Indices de la capacité prédictive du modèle dans le temps

La prédiction d'intensité d'observation de vivants est surestimée par rapport au jeu de test (de 1150.10% en moyenne) et le RMSE semble faible (2.61 en moyenne) mais il représente un très gros pourcentage de la moyenne observée donc une grosse part de variabilité (1321.56%), la prédiction des vivants est très imprécise (Fig. 10).

La prédiction des carcasses semble est sous-estimée (-80.92%) et son RMSE (0.87 en moyenne) comparé à la moyenne d'observation par maille de 24 km² représente 211.84% de la variabilité (Fig. 10).

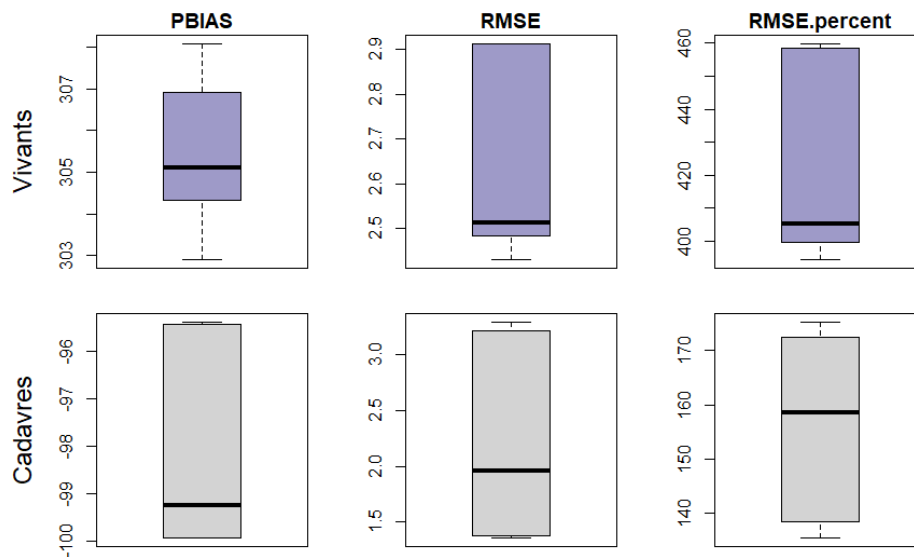


Figure 11 : Indices de la capacité prédictive spatiale du modèle

La prédiction d'intensité d'observation de vivants est surestimée par rapport au jeu de test (de 305.40% en moyenne) et le RMSE semble faible (2.62 en moyenne) mais et représente de variabilité (420.49% en moyenne) (Fig. 11). La prédiction des carcasses semble est sous-estimée (-98.19%) et son RMSE (2.19 en moyenne) comparé à la moyenne d'observation par maille de 24 km² représente 156.41% de la variabilité (Fig. 11).

4. DISCUSSION

Le modèle prédictif cartographiant le risque de mortalité routière du chevreuil à l'échelle de la Bretagne et des Pays de la Loire présente un ensemble de facteurs de complexité inhérent à l'utilisation de données opportunistes et à la modélisation d'événements rares. A ce stade du développement, son architecture est cohérente et son implémentation efficace, avec des paramètres biologiquement crédibles mais un ajustement et une capacité prédictive insatisfaisants. La couche d'exposition, qui se base sur les données opportunistes du programme Visionature, résulte en des prédictions fortement sous-estimées par rapport aux observations. En revanche, l'ajustement de la couche représentant le danger (modélisation des cadavres) est plus cohérente avec les observations, quoiqu'améliorable. Par conséquent, la cartographie prédictive du risque de mortalité ambitionnée en début de stage nécessite encore un travail de développement substantiel avant de pouvoir être proposée comme outil aux décideurs des politiques publiques d'aménagement du territoire. La cohérence biologique des paramètres suggère cependant que la source des problèmes à résoudre ne réside ni dans une spécification

inadéquate du modèle, ni dans une déficience des données. La discussion vise donc essentiellement à montrer comment les déficiences actuelles peuvent être dépassées.

Le modèle retranscrit de manière incomplète la variance induite par l'effort d'échantillonnage hétérogène dans la couche d'exposition. L'incorporation de cette variance additionnelle résulte de la modélisation a priori de l'effort d'échantillonnage basé sur une covariable spatiale (De Solan *et al.*, 2019; Johnston *et al.*, 2021), qui ne tient pas compte des variations individuelles de l'échantillonnage lié aux observateurs. Une espèce commune et de grande taille telle que le chevreuil est cependant peu susceptible de faire l'objet d'erreurs d'identification et, au moins en milieux ouverts, est habituellement bien détectée (Kelling *et al.*, 2015). En revanche, la résolution de la covariable d'échantillonnage est déterminante dans la qualité de sa modélisation, de fortes variations à des résolutions plus élevées peuvent être lissées par un maillage trop large. La résolution de la covariable d'échantillonnage choisie dans notre modèle (5*5km) pourrait être trop grossière pour rendre compte précisément de la variation de l'échantillonnage (Botella *et al.*, 2021) ; le grain spatial devra donc être exploré plus en détails dans les développements à venir. La qualité de la modélisation du danger (cadavres) fournit des résultats moins problématiques, même si l'intensité de cadavre prédite est biaisée en moyenne. On observe en effet des correspondances entre la prédiction du modèle et les observations sur certaines zones fortement mortifères pour les chevreuils, mais une incapacité du modèle à prédire les mortalités intermédiaires. Au regard de la structure jointe du modèle, une amélioration de la couche d'exposition devrait permettre d'améliorer également la qualité de l'ajustement de la couche de danger. L'intégration d'une composante d'autocorrélation spatiale peut informer le réalisme des prédictions du modèle, notamment en captant les autocorrélation issues de l'agrégation des animaux grégaires mais aussi de covariables d'influence indisponibles (présence de clôture, obstacle sur la route) (Bahn & McGill, 2007).

Conséquence d'un ajustement limité, le modèle présente une performance prédictive faible dans le temps comme dans l'espace. Outre l'ajustement du modèle, la qualité prédictive interannuelle est surtout limitée par la diminution du nombre de données notamment pour le jeu d'observation permettant de calibrer la couche d'exposition, rendant l'hétérogénéité de l'effort d'échantillonnage d'autant plus sensible (Johnston *et al.*, 2021). La prédiction sur un temps long des collisions animal-véhicule peut éventuellement bénéficier d'une incorporation de la variation temporelle du paysage, mais cette amélioration est peu susceptible d'apporter des changements substantiels dans la mesure où les paysages bretons ont relativement peu changé sur le pas de temps considéré dans notre étude (Medinas *et al.*, 2021; Rendall *et al.*, 2021). La mauvaise

capacité prédictive du modèle dans l'espace provient plus probablement des structures d'autocorrélations spatiales des covariables et des observations.

Malgré les déficiences d'ajustement et de qualité prédictive, les effets estimés des covariables sont conformes aux connaissances actuelles sur l'écologie du chevreuil. Les plus fortes intensités de présence du chevreuil en forêt et prairie correspondent à ses habitats préférentiels, tandis que l'effet négatif des zones urbaines et des cultures transcrivent son aversion pour les zones anthropisées (Tellería & Virgós, 1997; Borowik *et al.*, 2013; Ciach & Fröhlich, 2019). Une augmentation de la vitesse est un prédicteur aggravant des collisions routières déjà soulevé dans la littérature (Farmer & Brooks, 2012), tandis que le trafic, source de nuisance pour la faune, peut être une source d'évitement lorsqu'il est trop élevé (Bruinderink & Hazebroek, 1996; Benítez-López *et al.*, 2010). Néanmoins, le trafic journalier moyen annuel n'est pas représentatif des variations de trafic intra-journalière, et notamment du trafic nocturne, période la plus mortifère pour les chevreuils (Bíl *et al.*, 2020; Ignatavičius *et al.*, 2020) – il est cependant peu probable que nous puissions disposer d'un tel niveau de finesse temporelle pour ces données. Les effets prédits de la proximité à la végétation et à l'hydrologie comme impact aggravant des collisions routières sont généralement causés par la présence d'habitat ou de corridor écologique proche de la route (Malo *et al.*, 2004; Girardet *et al.*, 2015). Une amélioration de l'intégration des prédicteurs de distribution et de collision peut être apportée en les modélisant par des effets non-linéaires permettant ainsi une compréhension plus fine de leurs impacts (Guisan & Zimmermann, 2000; Guisan *et al.*, 2017).

Le modèle développé dans ce rapport a pour perspective de pouvoir inférer sur le risque de collision routière en tout point de l'espace sur l'étendue de la Bretagne. L'objectif prioritaire des politiques territoriales sur des enjeux émergents comme les collisions faune-véhicules étant de réaliser un état des lieux initial cartographique, une correction de l'ajustement du modèle est prioritaire. L'amélioration de la capacité prédictive du modèle est un développement nécessaire pour la planification spatiale des actions de mitigation et d'évitement du risque, mais peut venir dans un second temps. Pour parvenir à des résultats opérationnels, crédibles aux yeux des gestionnaires d'infrastructures et de l'administration territoriale, le travail de développement doit se poursuivre avec comme priorité une amélioration de la prise en compte des hétérogénéités d'effort d'échantillonnage sur la couche d'exposition.

Si le modèle développé dans ce rapport nécessite encore un travail substantiel, il a mis lumière la possibilité technique de construire un modèle répondant à la demande publique (cartographie prédictive) à partir des données disponibles, non spécifiquement récoltées dans ce

but et non protocolé. En cela, l'outil INLA constitue un gain de temps de calcul indispensable à cette échelle spatiale, d'autant plus que les nécessaires explorations de la structure du modèle sont incompatibles avec les approches MCMC, bien que plus conventionnelles. Notre étude a de plus démontré la difficulté d'exploiter les données opportunistes et / ou protocolées de manière hétérogène pour répondre à des questions de gestion et de politiques publiques, et ce malgré la forte promotion de la collecte et de l'analyse de telles données via, notamment, le Système d'Information sur la Nature et les Paysages (Système d'information de l'inventaire du patrimoine naturel (SINP), 2018). Cela implique que, malgré la possibilité de réaliser le modèle envisagé avec de telles données, ce projet nécessite une collaboration renforcée et sur la durée entre statisticiens, écologues et gestionnaires, souvent mise en valeur (Magurran *et al.*, 2010; Decker *et al.*, 2012; Sterling *et al.*, 2017) mais dans les faits, difficile à mettre en place. Notre modèle bénéficie d'une telle collaboration à travers le réseau Cisstat et le cadre offert par l'outil de financement Ittecop, visant justement à mettre en relation des chercheurs, porteurs d'enjeux et représentants du ministère chargé de l'écologie. Il permet de trouver le meilleur compromis possible entre qualité et efficacité de modélisation tout en créant une relation de confiance indispensable à la construction d'un modèle complexe sur le temps long. Par conséquent, bien que notre travail soit actuellement à un stade préliminaire, il offre suffisamment de garantie de réussite pour conclure qu'une solution de modélisation devrait permettre de répondre à l'objectif fixé avec des données non protocolées, avec toutefois un coût de développement élevé qu'il sera nécessaire de prendre en compte pour répondre à des demandes similaires ultérieurement.

REFERENCE

- Accueil - data.bioloVISION.net.
- Bahn, V. & McGill, B.J. (2007) Can niche-based distribution models outperform spatial interpolation? *Global Ecology and Biogeography*, **16**, 733–742.
- Benítez-López, A., Alkemade, R. & Verweij, P.A. (2010) The impacts of roads and other infrastructure on mammal and bird populations: A meta-analysis. *Biological Conservation*, **143**, 1307–1316.
- Bergès, L., Roche, P. & Avon, C. (2010) Corridors écologiques et conservation de la biodiversité, intérêts et limites pour la mise en place de la Trame verte et bleue. *Sciences Eaux & Territoires*, **Numéro 3**, 34–39.
- Bíl, M., Andrášik, R. & Janoška, Z. (2013) Identification of hazardous road locations of traffic accidents by means of kernel density estimation and cluster significance evaluation. *Accident Analysis & Prevention*, **55**, 265–273.
- Bíl, M., Andrášik, R. & Sedoník, J. (2019) A detailed spatiotemporal analysis of traffic crash hotspots. *Applied Geography*, **107**, 82–90.
- Bíl, M., Kubeček, J. & Andrášik, R. (2020) Ungulate-vehicle collision risk and traffic volume on roads. *European Journal of Wildlife Research*, **66**, 59.
- Billon, L. (2018) Rapport d'analyse de la répartition des collisions faune/véhicule DIR Ouest – Données récoltées de 2014 à 2016. *UMS 2006 Patrimoine naturel, AFB, CNRS, MNHN*, 71.
- Billon, L., Sordello, R. & Touroult, J. (2015) Protocole de recensement des collisions entre la faune sauvage et les véhicules: proposition d'un socle commun | Trame verte et bleue. *Service du patrimoine naturel, Museum national d'Histoire naturelle, Paris. SPN*, 18.
- Borowik, T., Cornulier, T. & Jędrzejewska, B. (2013) Environmental factors shaping ungulate abundances in Poland. *Acta Theriologica*, **58**, 403–413.
- Botella, C., Joly, A., Bonnet, P., Munoz, F. & Monestiez, P. (2021) Jointly estimating spatial sampling effort and habitat suitability for multiple species from opportunistic presence-only data. *Methods in Ecology and Evolution*, **12**, 933–945.
- Bruinderink, G.W.T.A.G. & Hazebroek, E. (1996) Ungulate Traffic Collisions in Europe. *Conservation Biology*, **10**, 1059–1067.
- Chyn, K., Lin, T.-E., Wilkinson, D.P., Tracy, J.L., Lawing, A.M. & Fitzgerald, L.A. (2021) Fine-scale roadkill risk models: understanding the intersection of wildlife and roads. *Biodiversity and Conservation*, **30**, 139–164.
- Ciach, M. & Fröhlich, A. (2019) Ungulates in the city: light pollution and open habitats predict the probability of roe deer occurring in an urban environment. *Urban Ecosystems*, **22**, 513–523.
- Clevenger, A.P., Barrueto, M., Gunson, K.E., Caryl, F.M. & Ford, A.T. (2015) Context-dependent effects on spatial variation in deer-vehicle collisions. *Ecosphere*, **6**, art47.
- Clevenger, A.P., Chruszcz, B. & Gunson, K.E. (2003) Spatial patterns and factors influencing small vertebrate fauna road-kill aggregations. *Biological Conservation*, **109**, 15–26.
- De Solan, T., Renner, I., Cheylan, M., Geniez, P. & Barnagaud, J.-Y. (2019) Opportunistic records reveal Mediterranean reptiles' scale-dependent responses to anthropogenic land use. *Ecography*, **42**, 608–620.
- Decker, D.J., Riley, S.J. & Siemer, W.F. (2012) *Human Dimensions of Wildlife Management*, JHU Press.
- Duffett, D., D'Amico, M., Mulero-Pázmány, M. & González-Suárez, M. (2020) Species' traits as predictors of avoidance towards roads and traffic. *Ecological Indicators*, **115**, 106402.
- Escamilla Molgora, J.M., Sedda, L., Diggle, P. & Atkinson, P.M. (2022) A joint distribution framework to improve presence-only species distribution models by exploiting opportunistic surveys. *Journal of Biogeography*, **49**, 1176–1192.

- Fahrig, L. (2003) Effects of Habitat Fragmentation on Biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, **34**, 487–515.
- Fahrig, L. & Rytwinski, T. (2009) Effects of Roads on Animal Abundance: an Empirical Review and Synthesis. *Ecology and Society*, **14**.
- Farmer, R.G. & Brooks, R.J. (2012) Integrated risk factors for vertebrate roadkill in southern Ontario. *The Journal of Wildlife Management*, **76**, 1215–1224.
- Favilli, F., Bíl, M., Sedoník, J., Andrášik, R., Kasal, P., Agreiter, A. & Streifeneder, T. (2018) Application of KDE+ software to identify collective risk hotspots of ungulate-vehicle collisions in South Tyrol, Northern Italy. *European Journal of Wildlife Research*, **64**, 59.
- Girardet, X., Conruyt-Rogéon, G. & Foltête, J.-C. (2015) Does regional landscape connectivity influence the location of roe deer roadkill hotspots? *European Journal of Wildlife Research*, **61**, 731–742.
- González-Suárez, M., Zanchetta Ferreira, F. & Grilo, C. (2018) Spatial and species-level predictions of road mortality risk using trait data. *Global Ecology and Biogeography*, **27**, 1093–1105.
- van der Grift, E.A., van der Ree, R., Fahrig, L., Findlay, S., Houlahan, J., Jaeger, J.A.G., Klar, N., Madriñan, L.F. & Olson, L. (2013) Evaluating the effectiveness of road mitigation measures. *Biodiversity and Conservation*, **22**, 425–448.
- Grilo, C., Borda-de-Água, L., Beja, P., Goolsby, E., Soanes, K., le Roux, A., Koroleva, E., Ferreira, F.Z., Gagné, S.A., Wang, Y. & González-Suárez, M. (2021) Conservation threats from roadkill in the global road network. *Global Ecology and Biogeography*, **30**, 2200–2210.
- Guillera-Arroita, G., Morgan, B.J.T., Ridout, M.S. & Linkie, M. (2011) Species Occupancy Modeling for Detection Data Collected Along a Transect. *Journal of Agricultural, Biological, and Environmental Statistics*, **16**, 301–317.
- Guisan, A., Thuiller, W. & Zimmermann, N.E. (2017) *Habitat Suitability and Distribution Models: with Applications in R*, Cambridge University Press.
- Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.
- Ignatavičius, G., Ulevičius, A., Valskys, V., Trakimas, G., Galinskaitė, L. & Busher, P.E. (2020) Temporal patterns of ungulate-vehicle collisions in a sparsely populated country. *European Journal of Wildlife Research*, **66**, 58.
- Jackson, N.D. & Fahrig, L. (2011) Relative effects of road mortality and decreased connectivity on population genetic diversity. *Biological Conservation*, **144**, 3143–3148.
- Jasińska, K.D., Żmihorski, M., Krauze-Gryz, D., Kotowska, D., Werka, J., Piotrowska, D. & Pärt, T. (2019) Linking habitat composition, local population densities and traffic characteristics to spatial patterns of ungulate-train collisions. *Journal of Applied Ecology*, **56**, 2630–2640.
- Jetz, W., Carbone, C., Fulford, J. & Brown, J.H. (2004) The Scaling of Animal Space Use. *Science*, **306**, 266–268.
- Johnston, A., Hochachka, W.M., Strimas-Mackey, M.E., Ruiz Gutierrez, V., Robinson, O.J., Miller, E.T., Auer, T., Kelling, S.T. & Fink, D. (2021) Analytical guidelines to increase the value of community science data: An example using eBird data to estimate species distributions. *Diversity and Distributions*, **27**, 1265–1277.
- Kaufman, L. & Rousseeuw, P.J. (2009) *Finding Groups in Data: An Introduction to Cluster Analysis*, John Wiley & Sons.
- Kelling, S., Johnston, A., Hochachka, W.M., Iliff, M., Fink, D., Gerbracht, J., Lagoze, C., Sorte, F.A.L., Moore, T., Wiggins, A., Wong, W.-K., Wood, C. & Yu, J. (2015) Can Observation Skills of Citizen Scientists Be Estimated Using Species Accumulation Curves? *PLOS ONE*, **10**, e0139600.
- Laurance, W.F. & Balmford, A. (2013) A global map for road building. *Nature*, **495**, 308–309.

- Linnell, J.D.C., Cretois, B., Nilsen, E.B., Rolandsen, C.M., Solberg, E.J., Veiberg, V., Kaczensky, P., Van Moorter, B., Panzacchi, M., Rauset, G.R. & Kaltenborn, B. (2020) The challenges and opportunities of coexisting with wild ungulates in the human-dominated landscapes of Europe's Anthropocene. *Biological Conservation*, **244**, 108500.
- MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Andrew Royle, J. & Langtimm, C.A. (2002) Estimating Site Occupancy Rates When Detection Probabilities Are Less Than One. *Ecology*, **83**, 2248–2255.
- Magurran, A.E., Baillie, S.R., Buckland, S.T., Dick, J.McP., Elston, D.A., Scott, E.M., Smith, R.I., Somerfield, P.J. & Watt, A.D. (2010) Long-term datasets in biodiversity research and monitoring: assessing change in ecological communities through time. *Trends in Ecology & Evolution*, **25**, 574–582.
- Malo, J.E., Suárez, F. & Díez, A. (2004) Can we mitigate animal–vehicle accidents using predictive models? *Journal of Applied Ecology*, **41**, 701–710.
- Medinas, D., Marques, J.T., Costa, P., Santos, S., Rebelo, H., Barbosa, A.M. & Mira, A. (2021) Spatiotemporal persistence of bat roadkill hotspots in response to dynamics of habitat suitability and activity patterns. *Journal of Environmental Management*, **277**, 111412.
- Medrano-Vizcaíno, P., Grilo, C., Silva Pinto, F.A., Carvalho, W.D., Melinski, R.D., Schultz, E.D. & González-Suárez, M. Roadkill patterns in Latin American birds and mammals. *Global Ecology and Biogeography*, **n/a**.
- Nos atlas régionaux - Bretagne Vivante.
- Opitz, T. (2017) Latent Gaussian modeling and INLA: A review with focus on space-time applications. *Journal de la société française de statistique*, **158**, 62–85.
- Papazoglou, I.A., Aneziris, O.N., Bellamy, L.J., Ale, B.J.M. & Oh, J. (2017) Quantitative occupational risk model: Single hazard. *Reliability Engineering & System Safety*, **160**, 162–173.
- Polak, T., Rhodes, J.R., Jones, D. & Possingham, H.P. (2014) Optimal planning for mitigating the impacts of roads on wildlife. *Journal of Applied Ecology*, **51**, 726–734.
- Raymond, S., Schwartz, A.L.W., Thomas, R.J., Chadwick, E. & Perkins, S.E. (2021) Temporal patterns of wildlife roadkill in the UK. *PLOS ONE*, **16**, e0258083.
- Rendall, A.R., Webb, V., Sutherland, D.R., White, J.G., Renwick, L. & Cooke, R. (2021) Where wildlife and traffic collide: Roadkill rates change through time in a wildlife-tourism hotspot. *Global Ecology and Conservation*, **27**, e01530.
- Renner, I.W., Elith, J., Baddeley, A., Fithian, W., Hastie, T., Phillips, S.J., Popovic, G. & Warton, D.I. (2015) Point process models for presence-only analysis. *Methods in Ecology and Evolution*, **6**, 366–379.
- Renner, I.W. & Warton, D.I. (2013) Equivalence of MAXENT and Poisson Point Process Models for Species Distribution Modeling in Ecology. *Biometrics*, **69**, 274–281.
- Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J.J., Schröder, B., Thuiller, W., Warton, D.I., Wintle, B.A., Hartig, F. & Dormann, C.F. (2017) Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, **40**, 913–929.
- Rue, H., Martino, S. & Chopin, N. (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **71**, 319–392.
- Santos, S.M., Lourenço, R., Mira, A. & Beja, P. (2013) Relative Effects of Road Risk, Habitat Suitability, and Connectivity on Wildlife Roadkills: The Case of Tawny Owls (*Strix aluco*). *PLOS ONE*, **8**, e79967.
- Sterling, E.J., Betley, E., Sigouin, A., Gomez, A., Toomey, A., Cullman, G., Malone, C., Pekor, A., Arengo, F., Blair, M., Filardi, C., Landrigan, K. & Porzecanski, A.L. (2017) Assessing the evidence for stakeholder engagement in biodiversity conservation. *Biological Conservation*, **209**, 159–171.

Système d'information de l'inventaire du patrimoine naturel (SINP) (2018) *PatriNat, centre d'expertise et de données sur le patrimoine naturel*.

Tellería, J.L. & Virgós, E. (1997) Distribution of an increasing roe deer population in a fragmented Mediterranean landscape. *Ecography*, **20**, 247–252.

Trame verte et bleue *Ministères Écologie Énergie Territoires*.

Visionature *Faune Sauvage*.

Warton, D.I. & Shepherd, L.C. (2010) Poisson point process models solve the “pseudo-absence problem” for presence-only data in ecology. *The Annals of Applied Statistics*, **4**, 1383–1402.

Zimmermann Teixeira, F., Kindel, A., Hartz, S.M., Mitchell, S. & Fahrig, L. (2017) When road-kill hotspots do not indicate the best sites for road-kill mitigation. *Journal of Applied Ecology*, **54**, 1544–1551.

REMERCIEMENT

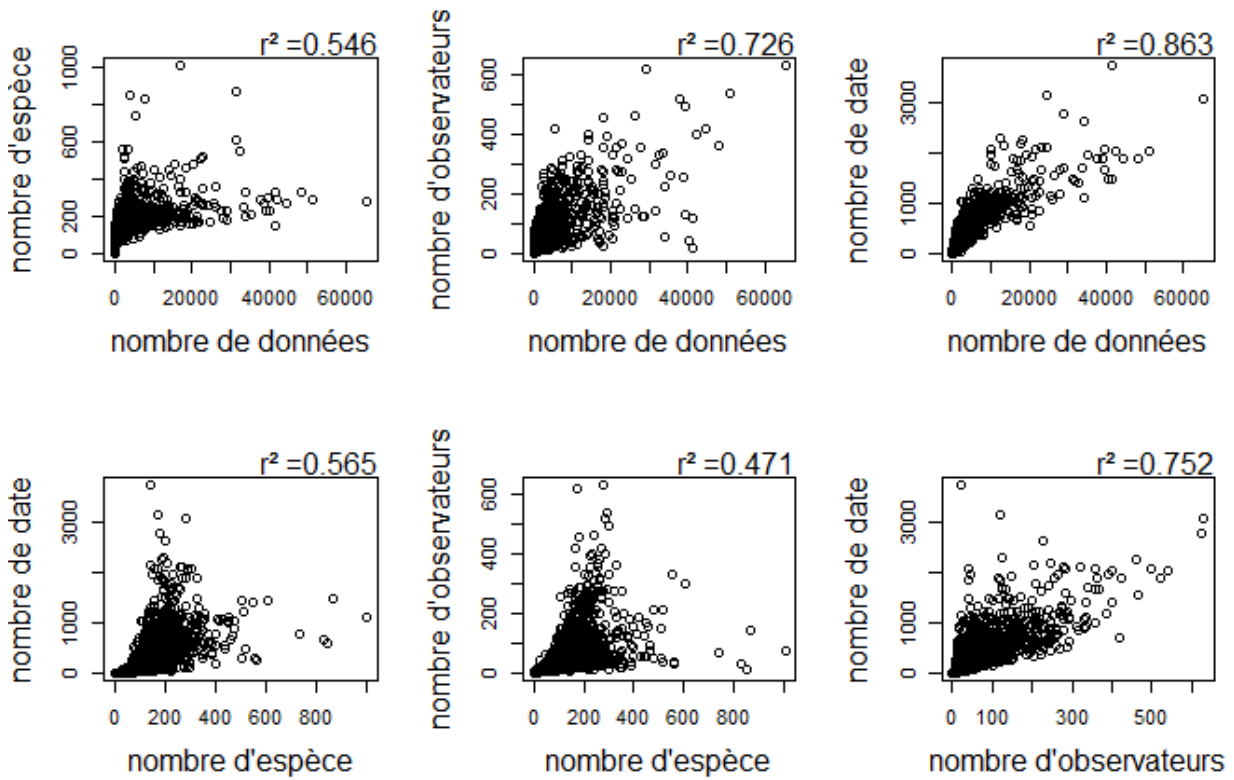
Je tiens à remercier le Cerema qui a financé ce projet ainsi que mes encadrants de stage, Jean-Yves Barnagaud, Jean-François Bretaud et Alain Morand pour leur accueil, leurs conseils, et les nombreux échanges qui ont ponctué ce stage.

Je remercie grandement la LPO Anjou, Faune Bretagne et le Groupe Mammalogique Breton pour nous avoir fourni les données nécessaires à la réalisation d'un tel projet et pour leur disponibilité pour les éventuelles questions, notamment Benoît Marchadour et Franck Simonnet.

Je remercie chaleureusement l'équipe de l'INRAE BioSP à Avignon qui m'a accueillie l'espace de quelques jours et notamment Julien Papaix qui m'a énormément aidée à la mise en forme ainsi qu'à la compréhension du modèle ainsi que ses explications sur les statistiques spatiales.

Également, un grand merci à l'équipe de Biologie Comportementale du Cefe à Montpellier pour m'avoir accueillie durant la durée de mon stage et pour le cadre de travail qu'ils m'ont fourni.

ANNEXE : Les covariables (cartes et corrélation)



| | Axe 1 | Axe 2 |
|-------------------------|-------|-------|
| Nombre de données | -0.92 | -0.15 |
| Nombre d'espèces | -0.72 | 0.69 |
| Nombre d'observateur | -0.86 | -0.27 |
| Nombre de dates uniques | -0.93 | -0.14 |

