



# Collisions Opportunistes et Collisions Protocoles liées aux Infrastructures de Transport Terrestres

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

JANVIER 2023

Le Cerema est un établissement public sous la tutelle du ministère de la Transition écologique, présent partout en métropole et dans les Outre-mer grâce à ses 26 implantations et ses 2 400 agents. Détenteur d'une expertise nationale mutualisée, le Cerema accompagne l'État et les collectivités territoriales pour la transition écologique, l'adaptation au changement climatique et la cohésion des territoires par l'élaboration coopérative, le déploiement et l'évaluation de politiques publiques d'aménagement et de transport.

Doté d'un fort potentiel d'innovation et de recherche incarné notamment par son institut Carnot Clim'adapt, le Cerema agit dans 6 domaines d'activités : Expertise & ingénierie territoriale, Bâtiment, Mobilités, Infrastructures de transport, Environnement & Risques, Mer & Littoral.

**Site web : [www.cerema.fr](http://www.cerema.fr)**

#### **Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres**

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

### **Statut de communication de l'étude**

Les études réalisées par le Cerema sur sa subvention pour charge de service public sont par défaut indexées et accessibles sur le portail documentaire du Cerema. Toutefois, certaines études à caractère spécifique peuvent être en accès restreint ou confidentiel. Il est demandé de préciser ci-dessous le statut de communication de l'étude.

- Accès libre : document accessible au public sur internet
- Accès restreint : document accessible uniquement aux agents du Cerema
- Accès confidentiel : document non accessible

Cette étude est capitalisée sur la plateforme documentaire [CeremaDoc](https://doc.cerema.fr/depot-rapport.aspx), via le dépôt de document : <https://doc.cerema.fr/depot-rapport.aspx>

### **Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres**

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

# Rappel du contexte de réalisation et des objectifs

## Contexte

Réponse à l'appel à proposition CILB - ITTECOP -FRB « Infrastructures situées : écosystèmes, paysages et dynamiques des territoires » - Appel à propositions de recherche 2020 du programme ITTECOP (Infrastructures de Transports Terrestres, ECOsystèmes et Paysages) »

**Axe écosystèmes.** Connaissance, gestion et réponse des écosystèmes comportant des ILT dans les territoires concernés

**Axe infrastructures.** Transformation des ILT et de leurs abords : nouveaux usages, nouvelles mobilités, nouvelles demandes écologiques et sociales

## Objectifs et résultats attendus (extraits du projet soumis à l'appel d'offre 2020)

Au-delà de créer une synergie entre plusieurs types d'acteurs (Cerema, EPHE/CEFE, DIR, collectivités (CD, régions), UMS Patrinat, associations naturalistes (LPO, etc), FDC et FRC, ...) autour d'une même problématique à enjeu élevé et d'une valorisation relative d'un jeu de données par rapport à l'autre, les résultats directement attendus de ce travail sont :

1. Un test de la capacité des modèles hybrides à produire des inférences robustes sur les données de mortalité effectivement disponibles. Il s'agit des modèles hiérarchiques associant processus ponctuels (couverture d'une grande surface et d'intervalles de temps long à partir de données non protocolées) et structures de type occupancy modeling (traitement de sources d'erreurs liées à l'acquisition des données : détection, hétérogénéité inter-observateurs par exemple).
2. Une identification des déficits de données tant en termes de mortalité que de ses prédicteurs, incluant à la fois la difficulté d'acquisition de certains types de données à grain très fin, les défauts de couverture spatio-temporelle et l'impact des principaux problèmes de protocole non solubles par la modélisation
3. Une première cartographie de la mortalité pour la région étudiée, basée sur les prédictions d'un modèle et une inférence sur un ensemble de facteurs biologiques et environnementaux considérés comme prioritaires (carte de présence et/ou distribution des espèces, cartes d'occupation du sol, données sur les ILTe, profil et type de routes, état du trafic, ...)
4. A moyen et plus long terme, une première réflexion et capacité à définir les priorités de recherches pour un projet plus ambitieux sur le couplage données protocolées /non protocolées en vue d'une mise en œuvre efficace de la séquence ERC

Cette base de données permettra de massifier les données pour des analyses plus approfondies sur les causes des collisions en couplant davantage de paramètres à grain fin (tels que le profil transversal de l'infrastructure de transport, la structure de végétation des bermes, présence de glissières type GBA ou d'échappatoires...) actuellement non disponibles. Elle permettra de mesurer avec plus de fiabilité les variations du nombre des collisions et l'efficacité des mesures de réduction, dont celles concernant les mises en place et le suivi des passages faune, que permettra également le « Système d'information recensant les passages à faune » (SIPAF), projet du Ministère de la Transition écologique développé en partenariat avec le Cerema et le Parc National des Ecrins. Il est en cours de développement et s'appuie sur Géonature. . Il sera conçu afin d'être compatible avec le Système d'Information sur la Nature et les Paysages (SINP) et avec le dépôt légal de Biodiversité (<https://depot-legal-biodiversite.naturefrance.fr/>). Cela facilitera l'échange et la valorisation des données, notamment dans l'inventaire national du patrimoine naturel (INPN).

## Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

# Équipe projet (structure, personne, rôle dans le projet)

## Centre d'études et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement (Cerema)

### Cerema/Dter Est

- Alain Morand, Responsable d'études en environnement - biodiversité, aménagement et infrastructure (Chef de projet et référent de l'étude pour ITTECOP, co-encadrant du stage M2, membre du Comité de pilotage), contact : *alain.morand@cerema.fr*
- Luc Chrétien, Chef du Groupe Biodiversité, Aménagement et Nature en Ville - (Référént administratif de l'étude pour la Dter Est et ITTECOP)
- Maeliss Vandroux, Stagiaire de Master 2 (Université Toulouse III Paul Sabatier) du 01/02/2022 au 29/07/2022 puis vacataire du 14/11/2022 au 13/12/2022 sur financement du projet ITTECOP, contact : *maelissvandroux@gmail.com*

### Cerema/Dter Ouest

- Jean-François Bretaud, Chef de projet biodiversité - (Référént expert de la zone d'étude, co-encadrant principal du stage M2, membre du Comité de pilotage), contact : *jean-françois.bretau@cerema.fr*

### Cerema/Dter Haut-de-France

- Olivier Pichard, Responsable d'étude biodiversité et aménagement (membre du Comité de pilotage)

### Cerema/Dter Sud-Ouest

- Eric Guinard, Chef de projet / chargé de recherche (membre du Comité de pilotage)

## Ecole Pratique des Hautes Etudes (EPHE)

- Jean-Yves Barnagaud, Maître de conférence (HDR), de l'Ecole Pratique des Hautes Etudes dans l'unité mixte de recherche Centre d'Ecologie Fonctionnelle et Evolutive (CEFE) (Référént pour la supervision de la construction des modèles : socle du travail de stage et du projet ; encadrant principal du sujet de stage, membre du Comité de pilotage). *jean-yves.barnagaud@cefe.cnrs.fr*

## Réseau scientifique CISSTAT (Citizen Science Statistics)

- Julien Papaïx, Chercheur à l'INRAE Avignon (Référént en modélisation statistique pour le stage), *julien.papaix@inrae.fr*

## Associations naturalistes

- Benoît Marchadour. Chargé de mission LPO Pays de la Loire (membre du Comité de pilotage)
- Franck Simonnet. Chargé de mission Groupe Mammalogique Breton (membre du Comité de pilotage)
- Laurent Couzi. Chef de service LPO France (contribution aux réflexions initiales du projet)

## Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

## Ministère de la transition écologique (DIR Ouest)

- Astrid Thomas-Bourgneuf, Chargée de mission développement durable (Représentante du gestionnaire du réseau routier étudié, membre du Comité de pilotage)

## Citation du document

Équipe projet « COCPITT » (Vandroux, M et al.) 2023. Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres. Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes. Projet 2021-2022, Cerema/EPHE. Version finale de mars 2023. 36 p + annexes.

## REMERCIEMENTS

L'équipe-projet COCPITT tient à remercier l'ensemble de l'équipe animatrice du programme ITTECOP, en particulier Yannick Autret (CGEDD), ainsi que le Comité Scientifique, le Conseil d'Orientation, la Fondation pour la Recherche sur la Biodiversité, Nicolas Hette-Tronquart de l'Office Français pour la Biodiversité et Annabelle Cluzeau de la Direction générale des infrastructures de transports et de la mer (DGITM / Ministère de la transition écologique) pour le soutien financier, logistique et l'animation scientifique qui a permis la réalisation et la valorisation du projet.

En effet, les temps d'échanges, de la réunion de lancement aux points d'étape du stage M2 tout au long de l'année 2022, la journée de restitution ont favorisé le bon déroulement de cette étude.

## **Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres**

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

# Sommaire

<b>1</b>	<b>Territoires et acteurs.....</b>	<b>11</b>
1.1	Associations naturalistes.....	12
1.2	Direction Interdépartementale des Routes Ouest (DIR Ouest) .....	12
1.3	Communauté scientifique .....	12
<b>2</b>	<b>DONNÉES .....</b>	<b>13</b>
2.1	Données semi protocolées de la DIR Ouest .....	13
2.2	Données opportunistes Visionature.....	15
<b>3</b>	<b>CONSTRUCTION D'UN MODÈLE.....</b>	<b>17</b>
3.1	Modèle théorique exhaustif de la mortalité par collision routière .....	17
3.2	Simplification du modèle théorique.....	18
3.2.1	Suppression en vue de décomplexifier le modèle.....	18
3.2.2	Dimension spécifique .....	18
<b>4</b>	<b>MODÈLE RÉALISÉ : une analyse de risque .....</b>	<b>20</b>
4.1	Le modèle .....	20
4.2	Les prédicteurs .....	22
4.2.1	Effort d'échantillonnage .....	22
4.2.2	Prédicteurs de l'exposition (covariables de paysage) .....	24
4.2.3	Prédicteurs du danger (covariables de route) .....	24
4.3	L'implémentation du modèle.....	26
4.3.1	Equations .....	26
4.3.2	Résolution .....	28
4.3.3	Validation .....	28
<b>5</b>	<b>RÉSULTATS.....</b>	<b>29</b>
5.1	Description des données.....	29
5.2	Ajustement du modèle : pour les données Visionature de faune vivante.....	30
5.3	Ajustement du modèle : pour les données DIRO de collision .....	31
5.4	Estimation des paramètres .....	31
5.5	Cartographie préliminaire du risque de collision routière.....	33
<b>6</b>	<b>BILAN .....</b>	<b>34</b>
6.1	Discussion .....	34
6.1.1	Résumé des résultats.....	34
6.1.2	Questions autour des données opportunistes et de leur intégration.....	34
6.1.3	L'apport possible de la spatialisation du modèle.....	35
<b>7</b>	<b>Conclusion.....</b>	<b>36</b>

**8 Annexes .....37**

## Introduction

Les routes sont des infrastructures de transport clefs pour le dynamisme d'un territoire. Toutefois, par le maillage dense qu'elles forment, elles constituent des ruptures brutales de la connectivité écologique d'un paysage et sont à l'origine d'une mortalité par écrasement non négligeable de la faune (Fahrig, 2003; Jackson & Fahrig, 2011). Elles deviennent ainsi une zone de conflit régulière entre l'humain et la faune sauvage et un enjeu majeur pour les gestionnaires d'infrastructures et les acteurs de la conservation de la biodiversité. En effet, cette surmortalité, associée à la réduction des déplacements de la faune par effet barrière, représente une menace réelle pour le maintien de nombreuses espèces (Fahrig & Rytwinski, 2009; Benítez-López et al., 2010). Les conflits faune-véhicule portent également sur la sécurité routière puisque les accidents liés à la faune sauvage peuvent engendrer des dommages matériels et/ou corporels.

Face à ce double enjeu de conservation et de sécurité routière, les politiques publiques cherchent à réduire les collisions faune-véhicule tout en maintenant la connectivité écologique du territoire, par exemple à travers les trames vertes et bleues (Bergès et al., 2010). La construction de dispositifs permettant à la faune de traverser sur ou sous la route (écoduc, banquette, encorbellement), associés à des dispositifs de retenue (clôture, mur, panneau acoustique), permettent d'atténuer durablement l'impact des routes. (clôture, mur, panneau acoustique) (van der Griff et al., 2013; Polak et al., 2014, Nowicky et al 2022). Si le positionnement précis et la nature de tels aménagements au coût non négligeable se résolvent à l'échelle locale, la priorisation de leur construction est décidée davantage lors de planification à l'échelle régionale. Ainsi, il apparaît indispensable de disposer d'une cartographie du risque de conflits faune-véhicule à échelle régionale afin de fournir un outil d'aide à la décision aux politiques publiques.

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 accidentogènes 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 faune-véhicule (Favilli et al., 2018). Bien que ces méthodes descriptives soient efficaces pour situer les points noirs de mortalité, une approche par modélisation statistique, peut permettre de prédire les patrons spatiaux des collisions faune-véhicule par des variables choisies sous hypothèses biologiques, mais également de prédire ces collisions (Malo et al., 2004; Santos et al., 2013; Grilo et al., 2021).

De plus, dans une telle approche à étendue régionale, l'étude des collisions faune-véhicule peut être informée par la distribution spatiale des espèces. En effet l'occurrence de collision faune-véhicule n'est envisageable qu'à condition que la faune soit susceptible de traverser la route, en plus de dépendre du caractère dangereux de celle-ci (vitesse des véhicules, largeur de la chaussée à traverser, etc.). La cartographie du risque de collision faune-véhicule à étendue régionale nécessite donc la disponibilité de données sur les collisions mais également sur la présence de la faune, voir son abondance (augmentant la probabilité de rencontre entre les espèces et les véhicules). Cependant pour une aussi large étendue spatiale, il est trop coûteux de mettre en place des méthodes d'échantillonnage protocolées, qui sont lourdes à réaliser et dans le cas de l'échantillonnage sur les routes, dangereuses pour l'opérateur. Ces limites imposent donc d'utiliser des données non-protocolées ou semi-protocolées, comme celles récoltées par les observatoires participatifs de biodiversité (Billon et al., 2015). Cependant bien qu'abondantes et couvrant une étendue spatiale conséquente, ces données sont récoltées bénévolement par des observateurs volontaires. Elles sont donc soumises à de fortes variations dans d'échantillonnage qui est hétérogène en fonction des caractéristiques des observateurs (connaissance du terrain, préférences personnelles, fréquence 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). Outre l'apport des données opportunistes, l'étude des collisions faune-véhicule à échelle régionale bénéficie majoritairement des données semi-protocolées récoltées par les Directions Interdépartementales des Routes sur le réseau routier national non concédé français, celles-ci étant moins sujettes aux biais liés à la variation d'échantillonnage.

### **Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres**

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

Le projet COCPITT vise à cet effet à construire un modèle prédictif spatialisé de la mortalité de la faune sur les routes, à l'échelle d'un territoire, qui s'étend en régions Bretagne et Pays de la Loire. La construction de ce modèle se structure autour de deux axes :

1. La modélisation de la mortalité de la faune due aux collisions en intégrant la distribution de la faune sur le territoire ;
2. L'utilisation et le couplage de plusieurs sources de données semi-protocoles ou opportunistes en limitant l'effet des biais d'échantillonnage sur les prédictions du modèle.

# 1 TERRITOIRES ET ACTEURS

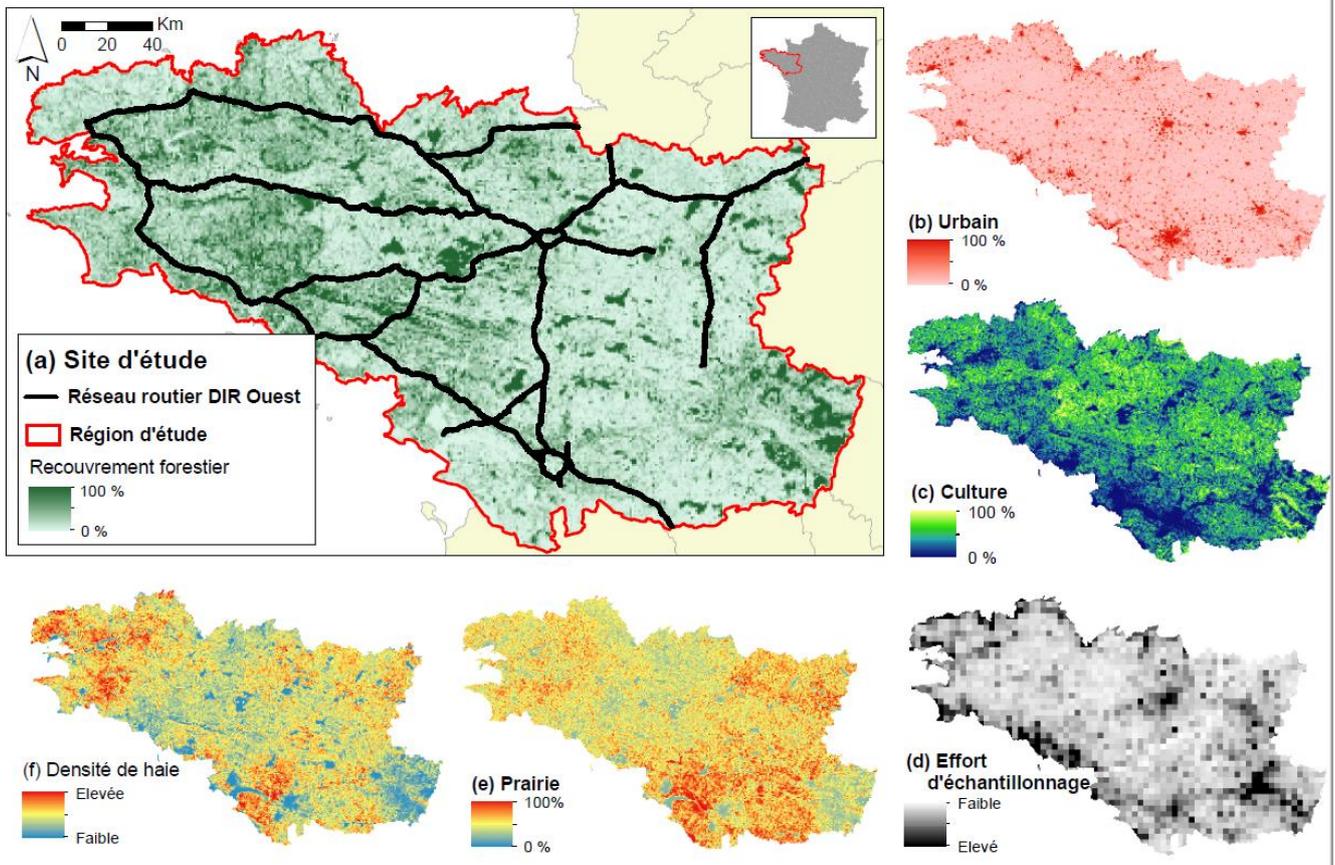


Figure 1: Site d'étude et son paysage : (a) Région d'acquisition des données, réseau routier de la DIR Ouest et cartographie du recouvrement forestier par km<sup>2</sup> (b) Recouvrement des zones urbaines par km<sup>2</sup> (c) Recouvrement des cultures par km<sup>2</sup> (d) Cartographie de l'effort d'échantillonnage par maille 5x5km (e) Recouvrement des prairies par km<sup>2</sup> (f) Densité de haie par km<sup>2</sup>

La région étudiée recouvre le territoire de la Bretagne et trois départements des Pays de la Loire (Loire Atlantique, Mayenne, Maine et 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 avoisinent les 400 mètres (Fig. 1). Ce territoire a été sélectionné notamment pour l'assiduité de la DIR Ouest vis-à-vis du protocole de relevé des données de collisions, et pour le dynamisme des associations naturalistes qui permet une couverture spatiale, temporelle et taxonomique élevées par des observations opportunistes.

La présente étude des collisions faune-véhicule s'inscrit dans une collaboration entre l'État (Ministère de la Transition Ecologique / DGITM), des experts de l'aménagement du territoire et de l'écologie de la route (Cerema) des chercheurs écologues modélisateurs, les associations naturalistes locales et la Direction Interdépartementale des Routes Ouest.

## Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

## 1.1 Associations naturalistes

Le réseau de science participative Visionature, initié par la LPO en 2007, rassemble les associations ornithologiques et naturalistes volontaires dans le but de combiner les données d'observations opportunistes à des fins de protection de la biodiversité. Les associations naturalistes locales sont les acteurs clefs de la mobilisation des naturalistes volontaires au sein des territoires de Bretagne et du Pays de la Loire et de la coordination de la base de données Visionature.

Les associations impliquées dans la récupération de données pour le territoire de Bretagne sont : la Ligue de Protection des Oiseaux (LPO), Bretagne Vivante-SEPNB et le Groupe Mammalogique Breton (F. Simonnet) .

Les associations impliquées dans la récupération de données pour le territoire du Pays de la Loire sont : la LPO (Anjou, Loire-Atlantique) (B. Marchadour), Mayenne nature environnement, Groupe Naturaliste de Loire Atlantique.

## 1.2 Direction Interdépartementale des Routes Ouest (DIR Ouest)

La Direction Interdépartementale des Routes (DIR) Ouest a mis en place depuis 2014 un protocole visant le recensement des collisions routières entre la faune et les véhicules qui rend disponible un jeu de données sur la mortalité entièrement indépendant des données de faune vivante. La DIR Ouest fait partie des DIR les plus assidues sur la récolte des collisions routières, stimulée par de nombreux partenariats et sessions 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 (80%) de 2x2 voies (Figure 1). La durée de la période étudiée débute en 2014, 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.

## 1.3 Communauté scientifique

Le développement du modèle des collisions routières a nécessité la mobilisation de la communauté de recherche scientifique. L'équipe du projet COCPITT se constituait en effet d'experts de l'aménagement du territoire (Cerema) mais également d'un chercheur de l'EPHE dont la recherche couvre notamment des questions liées à l'écologie du paysage et sa détermination de la distribution des espèces ainsi que les méthodes d'analyse de données pouvant répondre à ces questions (JY Barnagaud). Outre l'équipe COCPITT, le développement du modèle a été réalisé au cours d'un stage de 6 mois réalisé par une étudiante en modélisation des systèmes écologiques de l'université Toulouse III Paul Sabatier (M. Vandroux) et a demandé une forte mobilisation du réseau CiSStats (Citizen Science Statistics - J. Papaix). Ce réseau regroupe des statisticiens, des écologues modélisateurs et des porteurs d'enjeu afin de développer des méthodes statistiques aptes à intégrer et valoriser des jeux de données issues de sciences participatives.

## 2 DONNÉES

### 2.1 Données semi protocolées de la DIR Ouest

Les données de mortalité par collision sont récoltées quotidiennement par les agents DIR Ouest en parallèle 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 moyenne de 80 km/h.

De 2014 à 2021, les patrouilles ont détecté en moyenne 4620,75<sup>1</sup> carcasses par an sur 1590 km de route : ce qui correspondrait à 2,9<sup>2</sup> carcasses par kilomètre si les positions des collisions étaient réparties uniformément sur le réseau routier.

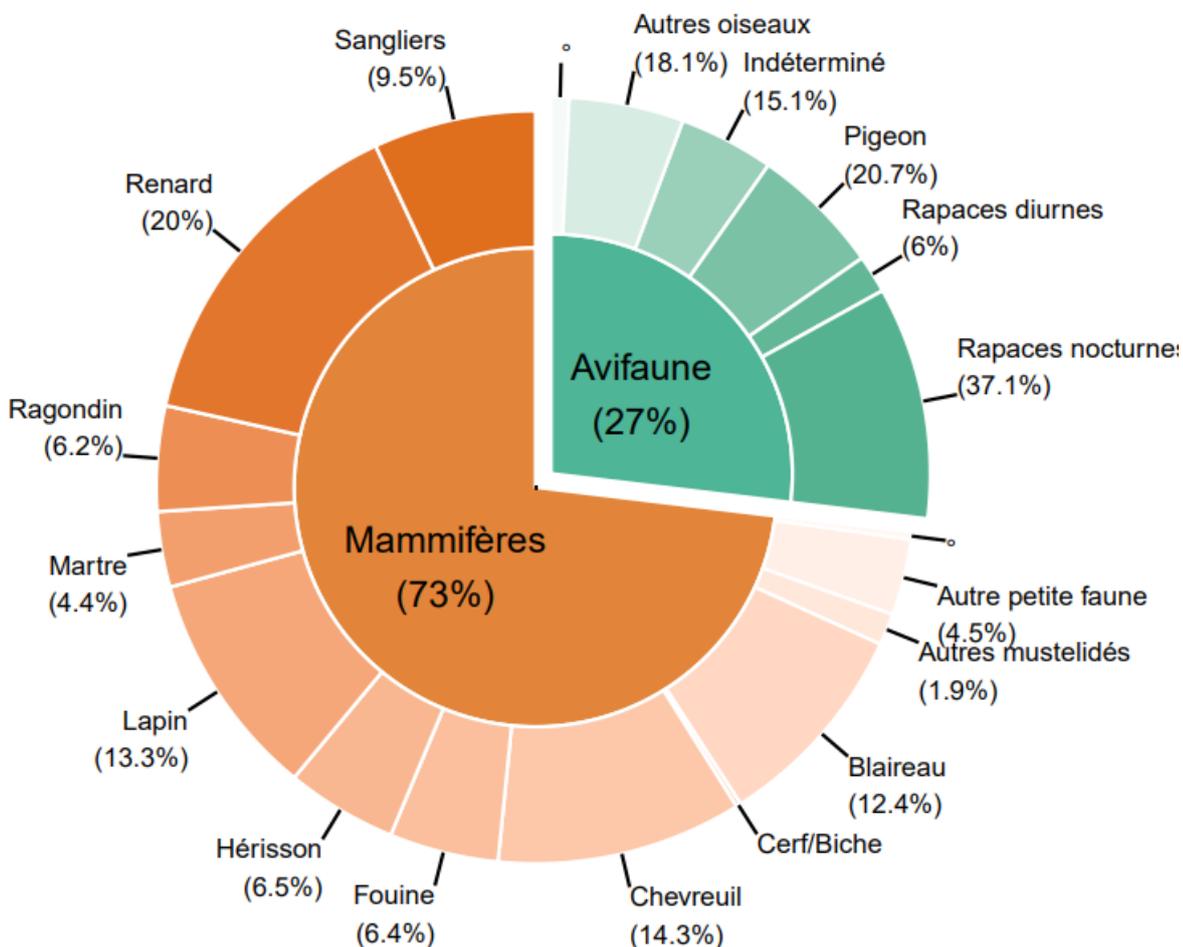


Figure 2: Composition des données de carcasse d'espèces sauvages observées par les agents de la DIR Ouest de 2014 à 2021 (à l'exception des amphibiens et des reptiles). Le symbole ° regroupe les espèces dont l'observation est anecdotique (<60 observations)

<sup>1</sup> Au minimum 3889 carcasses/an et au maximum 5232 carcasses/an

<sup>2</sup> Au minimum 2,45 carcasses/km/an et au maximum : 3,29 carcasses/km/an

#### Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

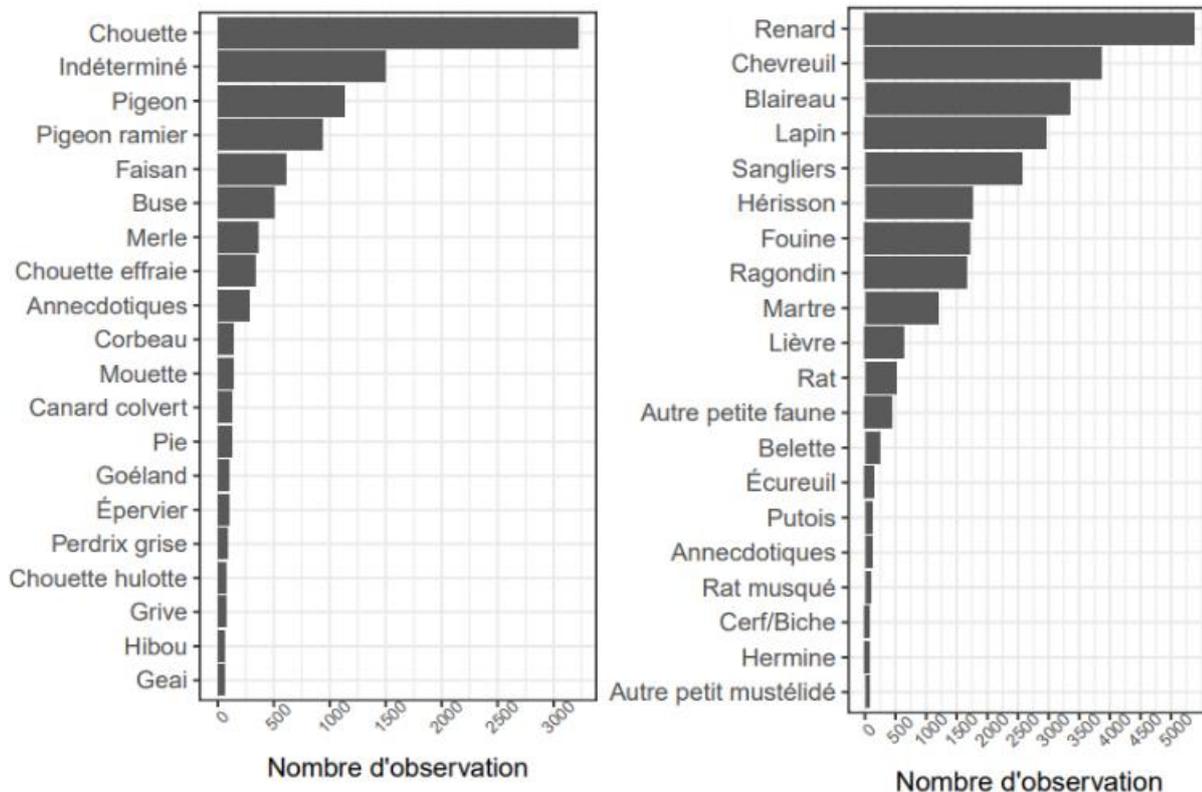


Figure 3: Occurrence de carcasses sur le réseau routier de la DIRO de 2014 à 2021

Les observations des patrouilleurs sont composées à 73 % d'observations de mammifères (~20 espèces) dont les carcasses appartiennent majoritairement à de la grande et moyenne faune : renard, chevreuil, blaireau, lapin, sanglier (Figure 2 et 3). Les oiseaux représentent quant à eux 27 % de la base de données de la DIR Ouest. Une cinquantaine d'espèces d'oiseaux a été identifiée suite à une collision routière mais seule une quinzaine de ces espèces concernent des détections très fréquentes (plus de 60 observations sur 8 ans). Les oiseaux principalement observés par la DIR Ouest sont les rapaces nocturnes, diurnes, les pigeons et les faisans. Les observations d'amphibiens (11 pour 8 ans) et de reptiles (22 pour 8 ans) restent très anecdotiques.

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 l'identification des carcasses est une tâche annexe aux missions des patrouilleurs de la DIRO, notamment celles de gestion et que ceux-ci ne sont pas spécifiquement formés à l'identification et ne sont pour la plupart pas naturalistes. En cas d'incohérence manifeste, les points de données ont été supprimés 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 ». Dans les cas où l'identification de l'espèce s'avère plus transparente malgré des erreurs de champs, les données ont été conservées avec les corrections nécessaires.

### Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

Afin d'avoir assez d'informations sur les collisions routières pour renseigner notre modèle, les espèces comportant un nombre suffisant d'observations de carcasses par les agents de la DIR Ouest ont été sélectionnées (plus de 60 carcasses au cours des 8 ans). 28 espèces<sup>3</sup> remplissent cette condition, ce sont les espèces pour lesquelles les données de faune vivante ont été fournies. Suite aux contraintes temporelles du projet, il a fallu retenir une espèce sur laquelle tester notre modèle.

Le chevreuil a été retenu 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). Une réflexion sur la modélisation multi-espèce a cependant été explorée et certaines tâches ont été réalisées à cet effet (se référer à la partie 3.2).

## 2.2 Données opportunistes Visionature

Les données de faune vivante proviennent de la base de données Visionature et ont été fournies par deux sources de compilation 'Pays de la Loire' et 'Faune Bretagne' impliquant quelques différences de champs disponibles selon les sources de compilation. Elles 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, les données opportunistes restent la source d'information sur les distributions d'espèces la plus fournie et accessible pour l'étendue spatiale visée (Figure 4). Les données opportunistes donnent accès à des résolutions temporelles 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.

Sont disponibles pour les années 2014 à 2021 : l'identification à l'espèce des espèces d'intérêt pour notre étude, les coordonnées de leur position, la date d'observation, l'identifiant de l'observateur, la précision de l'observation, l'état de l'animal (mort ou vivant). Seules les données à localisation précise ont été sélectionnées pour être utilisées dans le modèle. Les données concernant la faune vivante ont été sélectionnées pour estimer la distribution des espèces.

---

<sup>3</sup> Buse variable, Canard colvert, Chouette effraie, Chouette hulotte, Corbeau freux, Épervier d'Europe, Faisan de Colchide, Geai des chênes, Goéland argenté, Grive musicienne, Hibou moyen-duc, Merle noir, Mouette rieuse, Perdrix grise, Pie bavarde, Pigeon biset, Pigeon ramier, Belette d'Europe, Blaireau européen, Cerf élaphe, Chevreuil européen, Fouine, Hermine, Lièvre d'Europe, Martre des pins, Putois d'Europe, Renard roux, Sanglier

### **Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres**

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

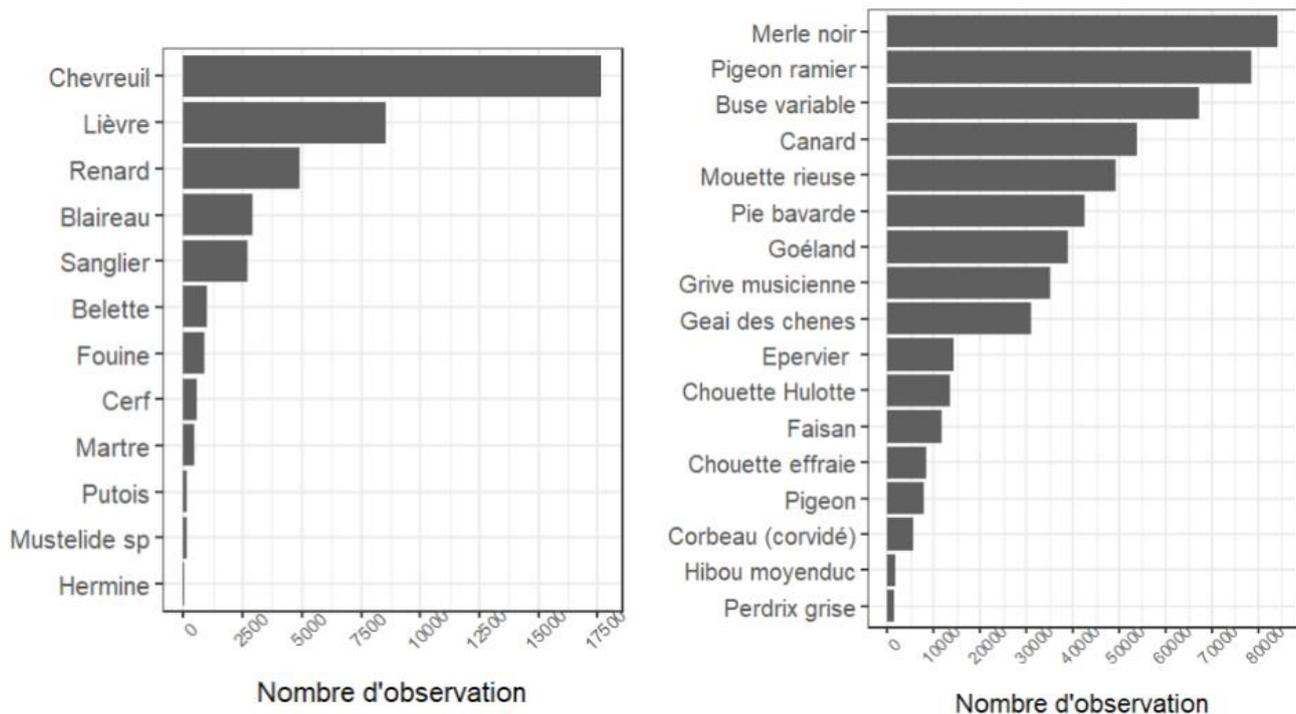


Figure 4 : Nombre d'observations d'animal vivant à une localisation précise de la base de données Visionature de 2014 à 2021 pour les Pays de la Loire et la Bretagne

L'application Naturalist utilisée pour renseigner les données de la base Visionature permet également de renseigner des informations sur l'état de l'animal et la cause probable de sa mort. Seules les données issues de la source de compilation 'Pays de la Loire' comportent le champ indiquant la cause probable de la mort qui permet de sélectionner uniquement les morts par collisions routières pour ce territoire. En revanche pour les données compilées par 'Faune Bretagne' fournies, s'il y a bien le champ renseignant l'état (mort ou vivant) de l'animal observé, aucun champ ne précise la cause de la mort. Pour ce territoire l'approche proposée pour sélectionner les observations d'animaux mort par collision routière, est de présumer qu'une carcasse est issue d'une collision routière si elle est située à moins de 15m des routes et/ou si l'observateur a utilisé des mots relatifs aux collisions routières dans la description de son observation (route, véhicule, écrasé, tapé, etc.). Les données opportunistes de collision routière n'ont pas pu être intégrées au modèle réalisé au cours du projet COCPITT à cause de la difficulté et du temps nécessaire à l'intégration de telles données. Les fichiers de données sélectionnant les données de collisions ne sont donc pas tous réalisés ni complets. Un script a cependant été réalisé permettant de sélectionner les morts par collisions présumées pour le territoire compilé par « Faune Bretagne ».

#### Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

### 3 CONSTRUCTION D'UN MODÈLE

#### 3.1 Modèle théorique exhaustif de la mortalité par collision routière

Le modèle a été construit selon trois dimensions, spatiale, temporelle et spécifique. La première, la dimension spatiale, qui vise à identifier les caractéristiques du paysage et de route qui déterminent l'agrégation spatiale des collisions routières à échelle régionale (Figure 5). Une collision est rendue possible à condition qu'un animal traverse la route, c'est-à-dire à condition qu'il soit présent sur le territoire. Les collisions routières sont donc dépendantes de l'abondance et la distribution des espèces au sein de la région étudiée. Cette abondance peut être par exemple estimée en modélisant l'habitat et la niche écologique des espèces à partir de caractéristiques de paysage (recouvrement en forêt, en zone urbaines, en cultures...). L'agrégation des collisions est également déterminée à l'échelle locale par les caractéristiques de la route qui la rendent plus ou moins dangereuse. Plusieurs caractéristiques de route ont été mises en évidence comme impactant le risque de collision : vitesse des véhicules, taille de la chaussée, topologie de la chaussée, présence d'un terre-plein arboré, présence d'obstacles (GBA), présence de clôtures pour la faune et de passage à faune, distance à la végétation, etc. (Benítez-López et al., 2010; Bíl et al., 2020).

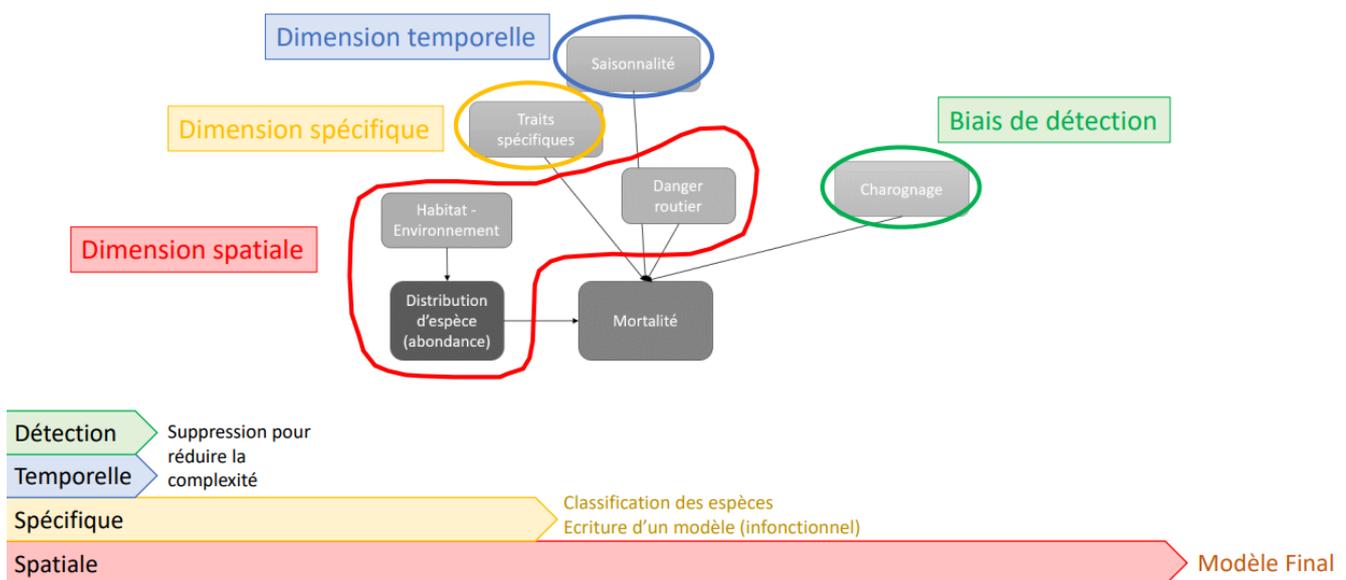


Figure 5 : Schéma du modèle théorique exhaustif de la mortalité de la faune par collision routière et frise synthétique des dimensions explorées dans le modèle du projet COCPITT

En plus d'être hétérogène dans l'espace, l'occurrence de collisions routières est hétérogène selon les saisons, c'est la dimension temporelle saisonnière. Le nombre de collisions ne se répartit pas de manière homogène tout au long de l'année mais est concentré sur certains mois plus mortifères que d'autres, ces périodes sont différentes selon les espèces étudiées. Cette saisonnalité est due à des variations dans le mouvement des animaux au sein du paysage, ces variations peuvent dépendre de la phénologie des espèces (saison de reproduction et mouvement migratoire), du comportement dispersif des stades de vie, notamment juvénile comme de phénomènes indépendants du cycle de vie des espèces. Ces derniers peuvent par exemple être liés au climat (augmentation des collisions avec les rongeurs pendant les périodes de crues) et à l'activité humaine (augmentation des collisions avec les sangliers lors des périodes de chasses). Les collisions routières avec la faune sauvage peuvent également être étudiées via une dimension temporelle intra-journalière, c'est-à-dire que l'occurrence de collision varie selon l'heure de la journée, et via une dimension temporelle à long terme qui considère l'évolution du nombre de collisions et de leur localisation sur plusieurs années (Ignatavičius et al., 2020; Raymond et al., 2021).

#### Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

La troisième dimension d'étude des collisions routières concerne les traits spécifiques. En effet la taille, le régime, le type de déplacement et le type d'habitat peuvent rendre une espèce plus ou moins vulnérable au danger routier (Medrano-Vizcaíno et al.). Il a par exemple été mis en avant qu'une espèce à régime charognard sera plus susceptible d'être victime d'une collision routière, car la nature de son régime l'attire vers les routes qui constituent une source de nourriture.

Le charognage constitue un biais de détection, l'un des plus importants, quasiment systématique de toute observation de carcasses d'animal sur une route. Il conduit à la disparition des carcasses sur la route, il peut intervenir par 3 processus. Le premier processus est la décomposition naturelle des cadavres dont la vitesse varie selon la taille de l'animal considéré et la météorologie (un temps humide augmente la vitesse de décomposition) (Santos et al., 2011). La vitesse de décomposition augmente en effet quand la taille et le poids de l'animal diminuent et elle est plus rapide chez les oiseaux que chez les mammifères. Les amphibiens et les reptiles ont une décomposition particulièrement rapide. Les deuxièmes et troisièmes processus sont le retrait des carcasses soit par des animaux charognard (Buse, renard, effraie...) soit par l'homme. (cf. Thèse en cours Annaëlle Bénard, Doctorante UMR CNRS 5023 - Laboratoire d'Ecologie des Hydrosystèmes Naturels et Anthropisés - Université Claude Bernard Lyon 1)

## 3.2 Simplification du modèle théorique

### 3.2.1 Suppression en vue de décomplexifier le modèle

Un modèle temporel, spatial et multi-spécifique est complexe et trop long à implémenter et tester sur 6 mois. Seules certaines parties du modèle ont donc été conservées dans le cadre du projet COCPITT. Les politiques territoriales étant demandeuses d'un outil d'aide à la décision, notamment pour la localisation d'aménagements pérennes (écoponts, encoirbellement...) il est nécessaire de conserver la dimension spatiale, cependant la dimension temporelle saisonnière n'est pas requise (les passages à faune pourront être traversés toute l'année). Le processus de charognage n'a pas été conservé, son développement aurait demandé de fortes connaissances sur la persistance de chaque espèce et une réflexion approfondie sur l'implémentation de cette persistance.

### 3.2.2 Dimension spécifique

La dimension spécifique a été explorée dans l'objectif de pouvoir tirer des conclusions sur des espèces pour lesquelles peu de données sont disponibles à partir d'informations sur le groupe dans lequel ont été classées ces espèces. La dimension spécifique du modèle a été mise de côté au cours du projet afin de perfectionner un modèle mono espèce plus simple à réaliser, cependant un travail de classification des espèces en groupes informatifs sur le risque de collision routière a déjà été mené.

Des traits écologiques ont été sélectionnés selon leur pertinence pour obtenir des groupes d'espèces ayant des facteurs de risques similaires vis-à-vis des routes (Annexe 2). Les traits utilisés pour la classification des espèces sont :

- La taille et le poids moyen de l'espèce (généralement corrélés à l'aire de déplacement des individus et déterminants pour la détection de la faune par les conducteurs) ;
- Le régime et la stratégie de recherche de nourriture de l'espèce (détermine sa mobilité et potentiellement son attrait pour les routes) ;
- Son grégarisme (impacte le nombre d'animaux qui traversent la route déterminant alors la capacité d'évitement de la part du conducteur) ;
- Son type de déplacement (vitesse pour les mammifères et type d'envol pour l'avifaune qui déterminent la capacité de l'animal à traverser rapidement la chaussée et/ou fuir à l'arrivée d'un véhicule) ;

#### **Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres**

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

- Son caractère nocturne ou non (impacte la détection par le conducteur).

Ces traits ont permis de calculer l'indice de similarité de Gower entre chaque couple d'espèces. Les espèces ont ensuite été regroupées par classification hiérarchique ascendante selon leur similarité grâce au critère d'agrégation du 'saut maximal' (Annexe 3). Les groupes ont été calculés séparément pour les mammifères et pour l'avifaune.

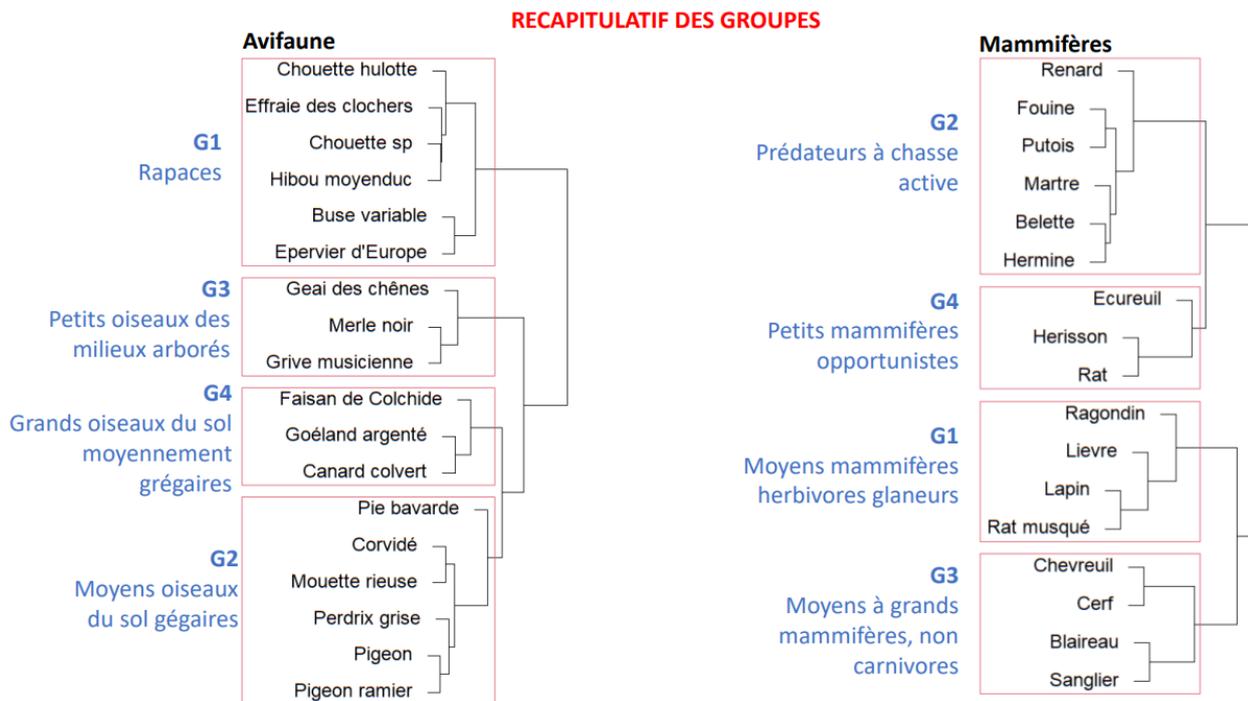


Figure 6 : Groupes d'espèces pouvant comporter des facteurs de risque de collision similaires, déterminés par classification hiérarchique ascendante

## 4 MODÈLE RÉALISÉ : UNE ANALYSE DE RISQUE

### 4.1 Le modèle

Le modèle spatial de mortalité de la faune sauvage par collision peut être analysé de manière analogue à une analyse de risque. Un risque est défini comme l'intersection entre un danger et une exposition (Figure 7). Dans le cas des collisions faune-véhicule, 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.7.a et Fig.7.b). Ce patron spatial correspond aux modèles statistiques de distribution 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. Le danger se comprend comme les propriétés de la route qui affectent la probabilité de collision à exposition donnée (Fig.7.b. et Fig.7.c).

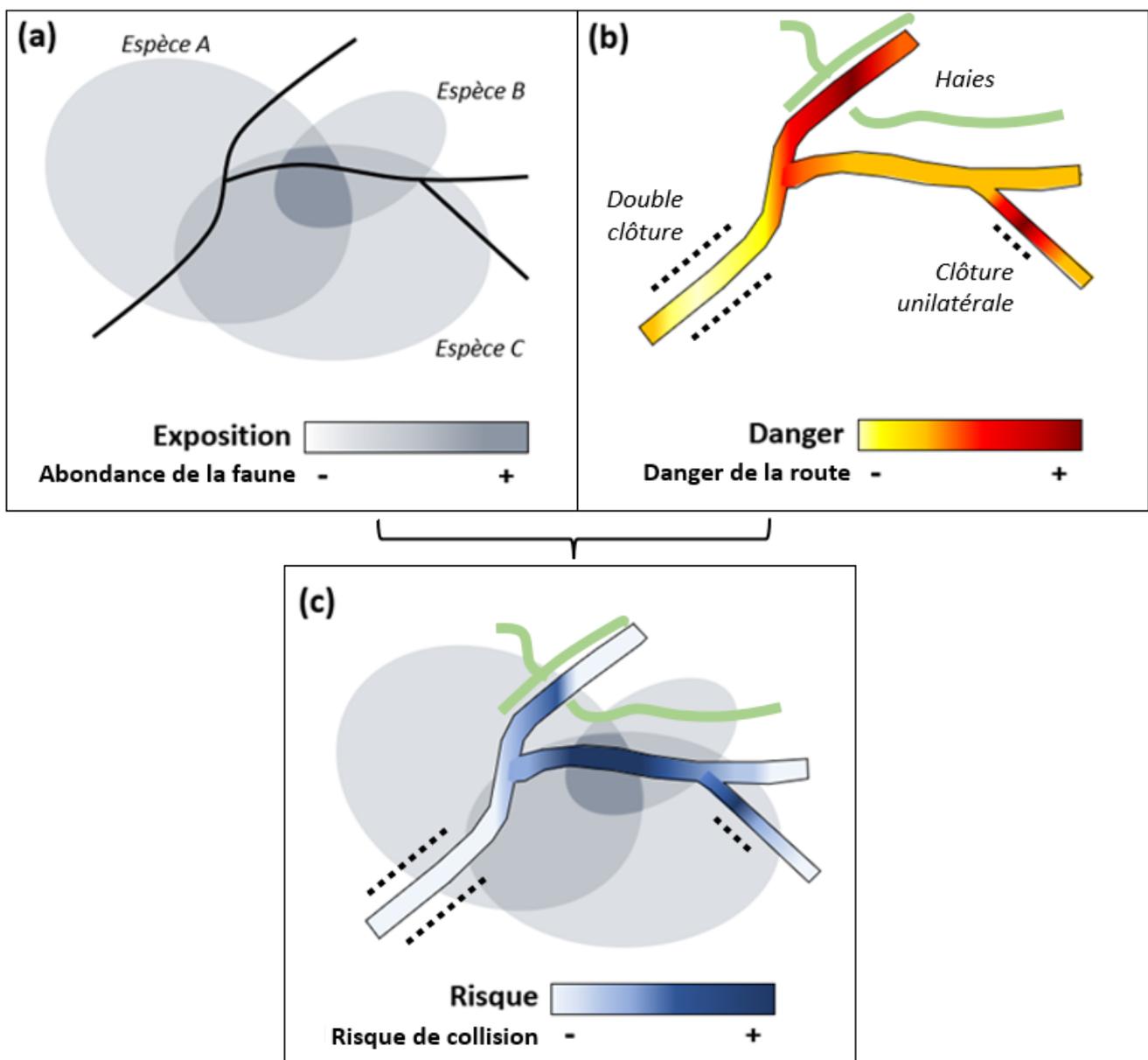


Figure 7: Schéma de principe d'une analyse de risque couplant l'exposition (a), placement d'un animal de manière qui l'expose aux effets de certains équipements, exprimée comme le recouvrement de la distribution d'une espèce avec un réseau routier, et le danger (b) propriété intrinsèque d'un équipement 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)

#### Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

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. 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) (González-Suárez et al., 2018).

Un modèle statistique nécessite d'être ajusté sur des données, ici il faut des données renseignant le danger routier mais également l'exposition (la distribution de l'espèce). Trois sources de données sont disponibles, pour estimer à la fois le danger et l'exposition : les données sur la faune morte par collision observée par la DIRO et celles observées par les utilisateurs de la base de donnée VISIONATURE ; enfin, pour séparer l'effet de l'exposition : les données sur la faune vivante de la base de donnée VISIONATURE.

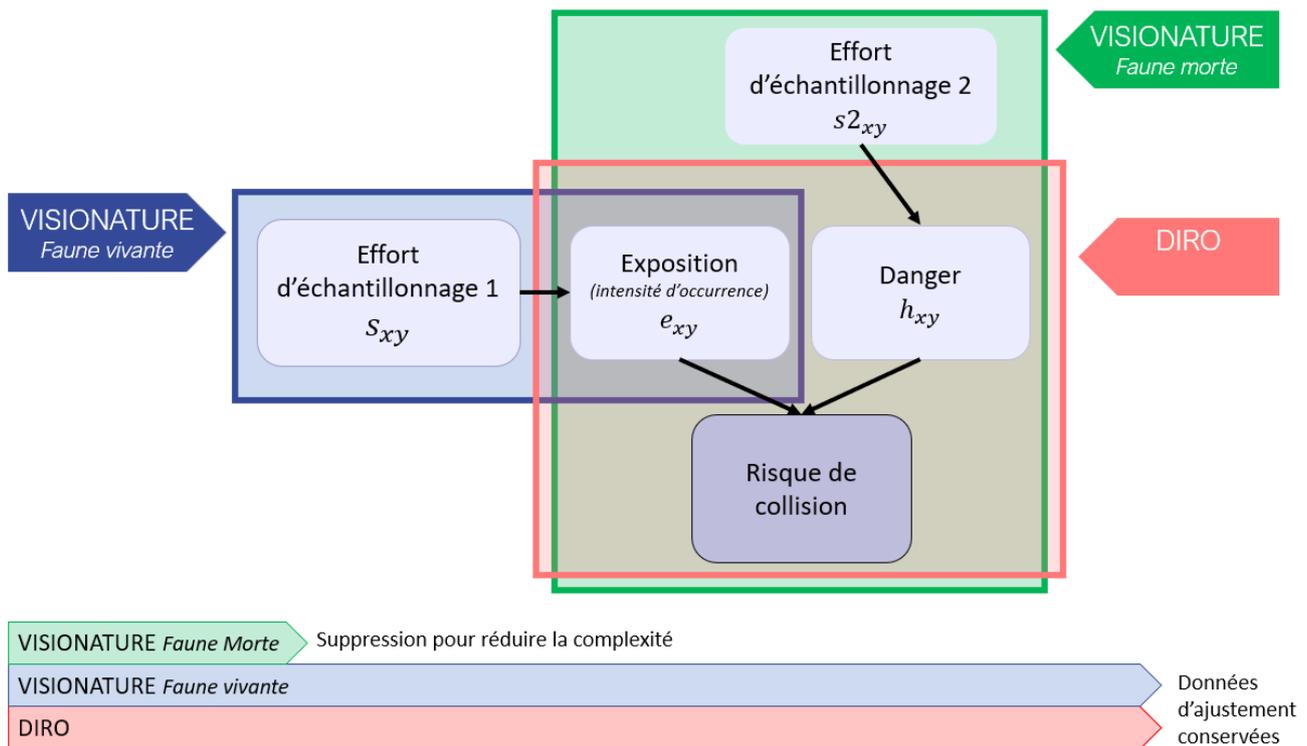


Figure 8 : Schéma des données d'ajustement du modèle et frise synthétique des données disponibles conservées dans le modèle du projet COCPITT (les SCRIPTS sont repris en annexe).

Les données issues de la base VISIONATURE étant opportunistes, il faut modéliser leur effort d'échantillonnage, cet effort d'échantillonnage est supposé différent entre l'échantillonnage de la faune vivante et l'échantillonnage de la faune morte par collision. Il est supposé que l'effort d'échantillonnage des données de la DIRO est constant et donc qu'il n'y a pas de nécessité de modéliser celui-ci. Pour diminuer la complexité du modèle, le nombre de sources de données a été restreint à deux. Les données Visionature de faune vivante ont été conservées pour bien estimer l'exposition. Et pour informer le danger, les données de la DIRO ont été conservées car elles bénéficient d'un effort d'échantillonnage homogène. En revanche, les données Visionature de la faune morte ont été écartées de l'analyse, car elles étaient moins nombreuses et nécessitaient le calcul de leur effort d'échantillonnage sur lequel peu d'informations étaient disponibles.

### Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres

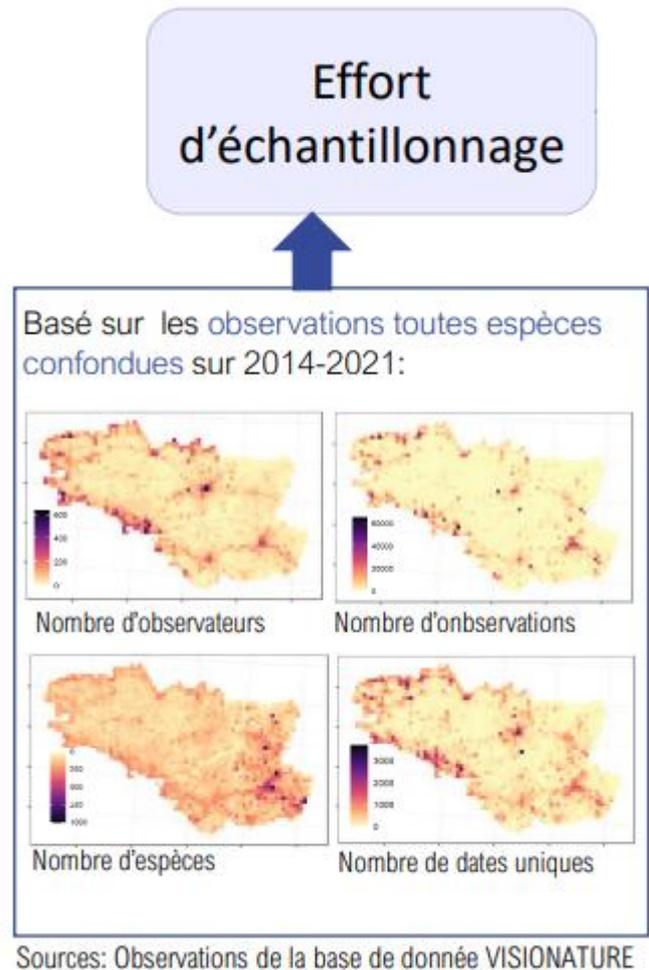
Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

## 4.2 Les prédicteurs

### 4.2.1 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). L'effort d'échantillonnage a été représenté à partir quatre variables synthétiques sur la base de données Visionature, sur une maille de 5 x 5 km et pour l'ensemble du territoire :

- Le nombre total de point d'observation (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 à 0.86 (Figure 9). Pour caractériser la colinéarité, une Analyse en Composantes Principales (ACP) centrée réduite a été réalisée 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).

#### Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

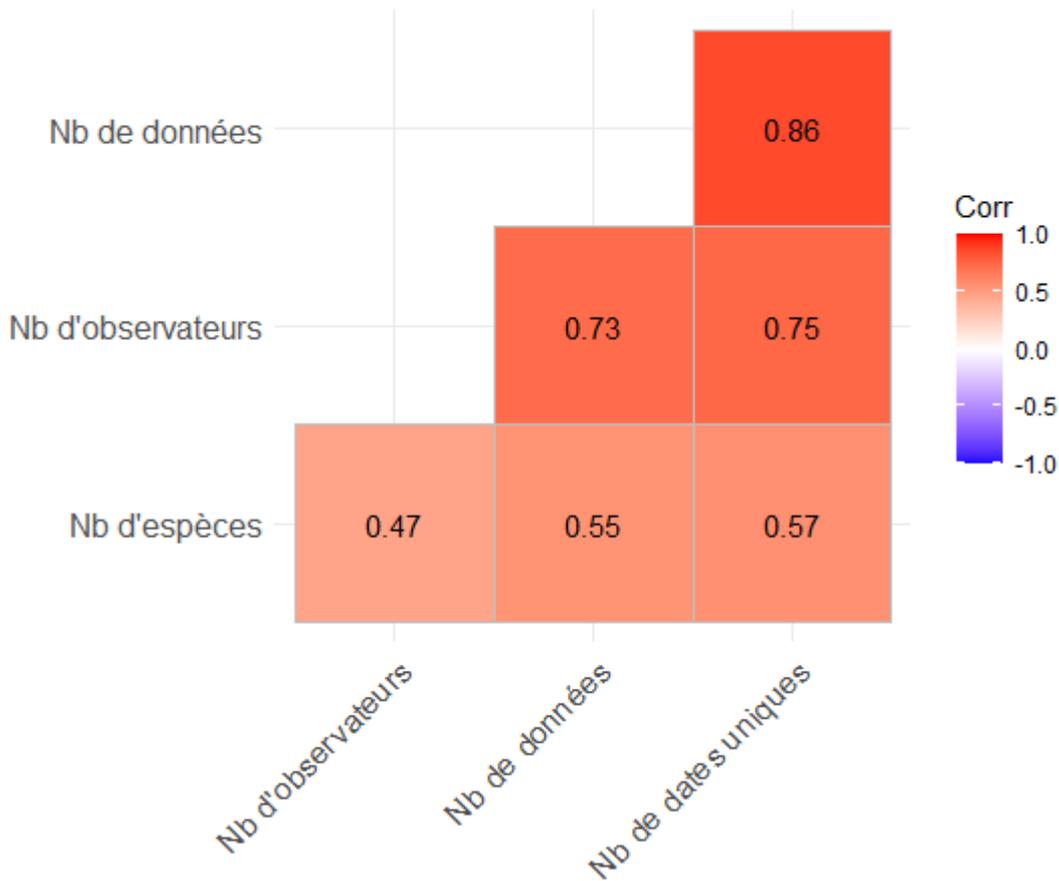


Figure 9 : Tableau des corrélations entre les variables de synthèse sur la base de données Visionature

Seul le premier axe a finalement été conservé (Figure 10). Pour des raisons de facilité d'interprétation, les valeurs de PC1 ont été inversées dans la covariable (eq. 1). Plus les valeurs de *sampling* sont élevées, plus l'effort d'échantillonnage est élevé.

$$sampling = -1 \times \text{Axe 1} \quad (\text{eq. 1})$$

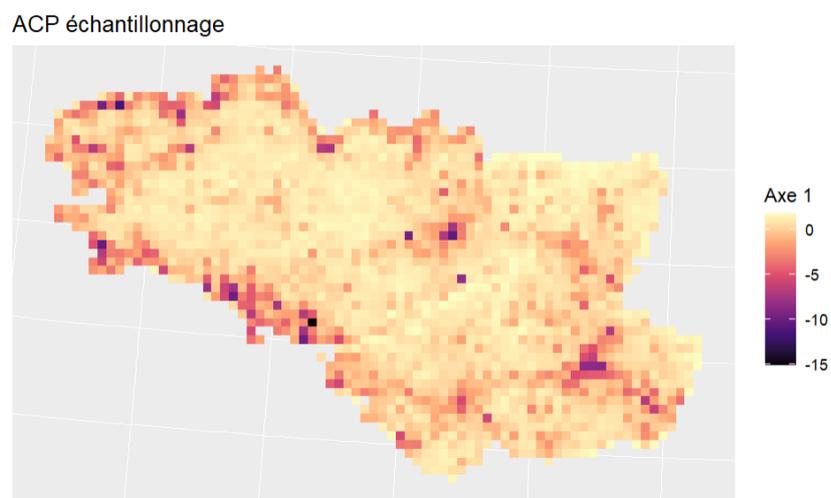


Figure 10 : Cartographie des scores du premier axe PC1 de l'ACP (les valeurs négatives correspondent à une forte pression d'échantillonnage)

### Collisions Opportunistes et Collisions Protocoles liées aux Infrastructures de Transport Terrestres

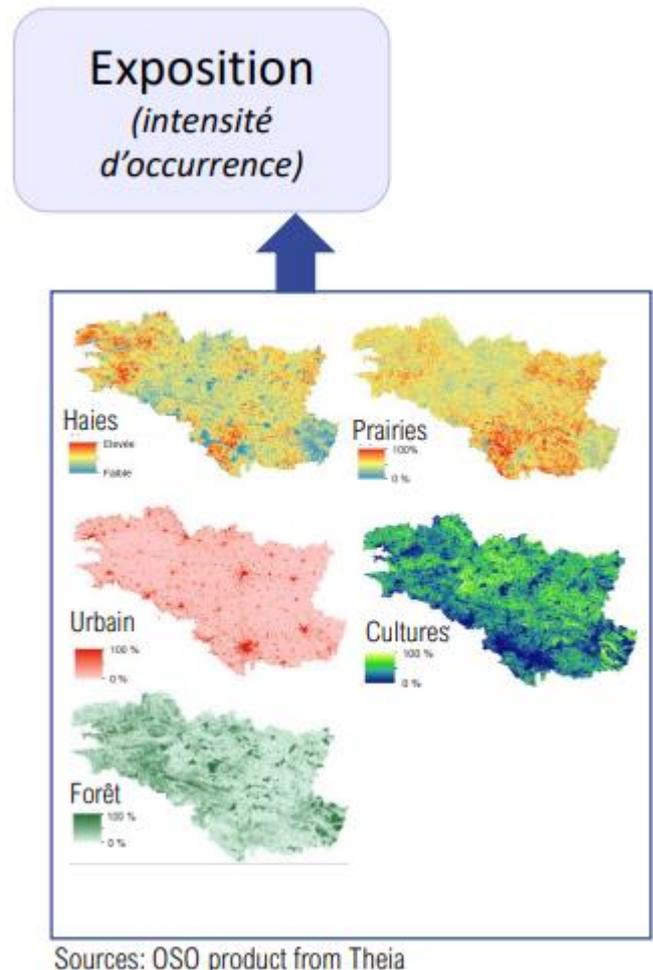
Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

Cette covariable comprend un nombre réduit de valeurs extrêmes. Afin d'équilibrer le poids des valeurs de la covariable *sampling*, une transformation racine carrée a été réalisée 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).

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

#### 4.2.2 Prédicteurs de l'exposition (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. 1). La densité de haie a été calculée par pixel à partir des fichiers de forme de 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), ont été calculés 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/>).



#### 4.2.3 Prédicteurs du danger (covariables de route)

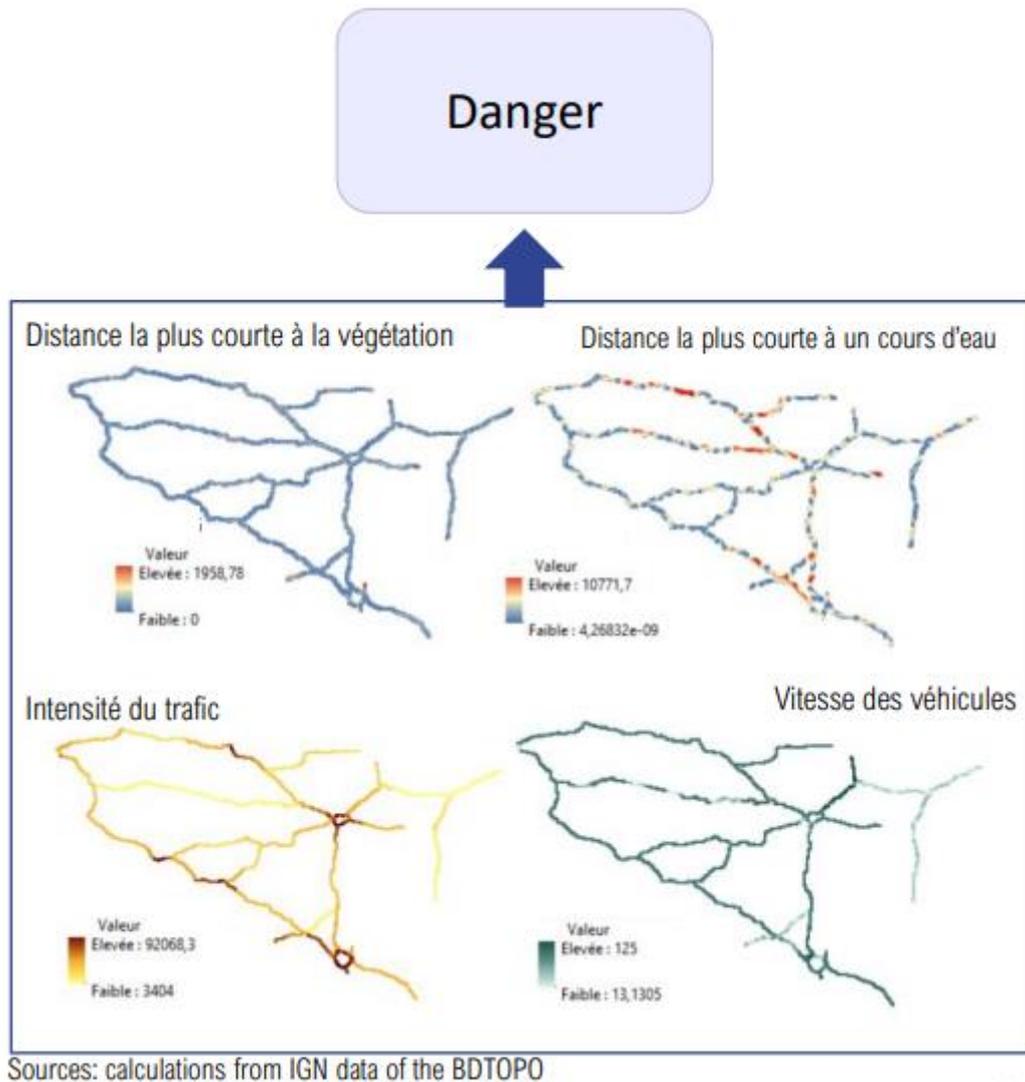
Les prédicteurs relatifs au danger routier sont les suivants :

- Intensité du trafic routier ;
- Vitesse des véhicules ;
- Proximité aux cours d'eau ;
- Proximité aux zones de végétation.

#### Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

Dans les paragraphes suivants, la source de données et les calculs effectués pour obtenir ces covariables est brièvement décrite.



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 traitement 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. Ces valeurs ont été moyennées (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 routes (autoroute, bretelle, route nationale...) et de leur importance pour le trafic routier (liaison entre grande métropoles, départements, grosses agglomérations, communes, etc.). La vitesse a été moyennée 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. La vitesse est donc considérée comme une variable qualitative à 2 classes : vitesse $\leq$ 90km/h nommée vitesse90 et vitesse $>$ 90km/h nommée vitesse100 .

Dans la présente étude, il est 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 au danger de la route pour la faune. Si cet environnement pourrait aussi être considéré comme une strate locale

#### **Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres**

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

de l'exposition, le choix a été fait de l'associer au danger afin de faciliter la construction du modèle. Les descripteurs de l'environnement local retenus sont la distance à la végétation la plus proche et la distance à un cours d'eau le plus proche. 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, les données sur l'hydrologie et la végétation ont été récupérées 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, il a été 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. Les cours d'eau majeurs ont été conservés (numéro d'ordre de 4 ou plus). Pour obtenir la donnée distance à ces cours d'eau, il a été 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.

### 4.3 L'implémentation du modèle

#### 4.3.1 Equations

Le risque de mortalité par collision routière des chevreuils a été analysé en modélisant conjointement les observations d'individus vivants ( $Y^{vivant}$ ) et les observations par la DIRO des carcasses sur la chaussée ( $Y^{collision}$ ) (Fig. 11)

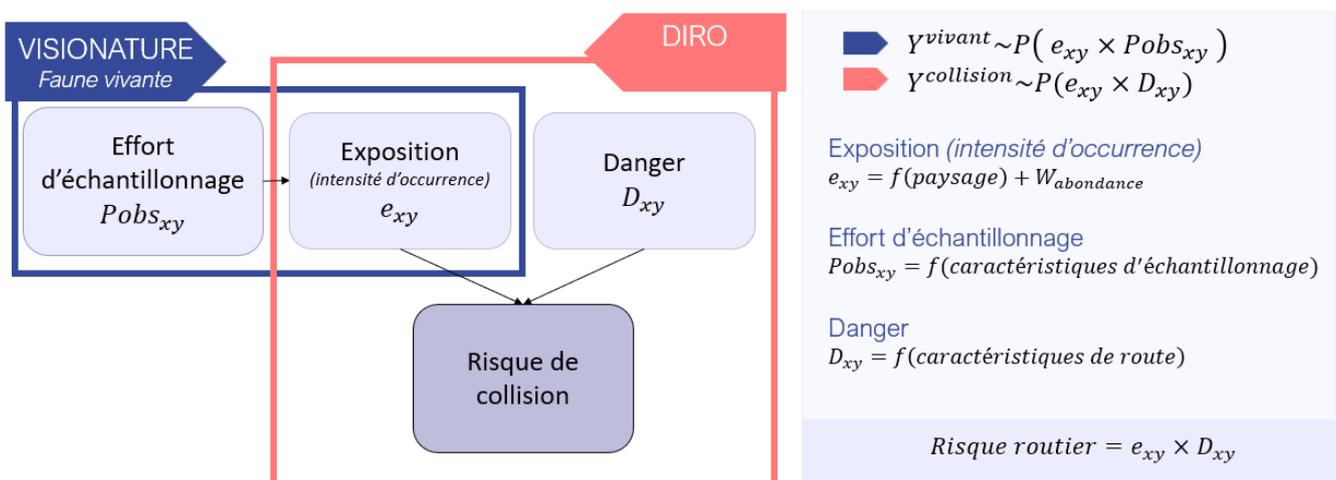


Figure 11 : Schéma du modèle réalisé (à gauche) et correspondance avec l'écriture mathématique du modèle (à droite)

#### Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

Ces deux jeux de données sont modélisés comme étant tirés chacun dans une loi de Poisson<sup>4</sup> inhomogène qui admet que l'intensité  $\lambda$  varie spatialement (eq. 3 et 4) (Renner et al., 2015).

$$Y_s^{vivant} \sim P(E_s * Pobs_s) \quad (\text{eq. 3})$$

$$Y_s^{collision} \sim P(E_s * D_s) \quad (\text{eq. 4})$$

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 la loi de Poisson inhomogènes des observations des vivants et observations des carcasses peuvent être décomposées 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. 3 et 4). C'est une modélisation de la distribution de l'espèce à partir de la niche écologique réalisée, exprimée par des prédicteurs paysagers (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 'Prédicteurs de l'exposition' partie 4.2 et eq 5). L'approximation  $E_s$  correspond à l'intensité d'occurrence de chevreuils qui peut s'interpréter comme une abondance relative de chevreuils.

$$\ln(E_s) = \alpha + \beta^{haie} \cdot HAI E_s + \beta^{culture} \cdot CULTURE E_s + \beta^{foret} \cdot FORET E_s + \beta^{prairie} \cdot PRAIRIE E_s + \beta^{urbain} \cdot URBAIN E_s \quad (\text{eq. 5})$$

La variable latente  $Pobs$  est la probabilité d'observer un chevreuil vivant (eq. 3). Elle est modélisée comme une régression logistique de la covariable de synthèse d'échantillonnage  $SAMPLING_s$  pondérée par un paramètre  $\beta^{sampling}$  (voir section 'Effort d'échantillonnage' partie 4.2 et eq 6). Le paramètre  $\beta^{sampling}$  a été forcé pour être supérieur ou égal à 0 puisque des valeurs élevées de la covariable  $SAMPLING_s$  induisent une forte pression d'échantillonnage et donc probabilité élevée d'observer un animal ; le paramètre  $\beta^{sampling}$  sert donc seulement à déterminer la forme de  $Pobs_s$  :

$$Pobs_s = \text{logit}(\beta^{sampling} \cdot SAMPLING_s) \quad (\text{eq. 6})$$

La variable latente  $D_s$  correspond au danger routier, soit à la variabilité de l'occurrence des carcasses de chevreuil sur les routes indépendamment de leur distribution (eq. 4). Le logarithme de  $D_s$  est modélisé à partir d'une combinaison de relation linéaires entre les covariables quantitatives de route, et à partir d'un intercept  $\alpha^{VITESS E_s}$  qui dépend de la classe de vitesse de la route à la localisation  $s$  (voir section 'Prédicteurs du danger' partie 4.2 eq.7).

$$\ln(D_s) = \alpha^{VITESS E_s} + \beta^{trafic} \cdot TRAFIC_s + \beta^{dist_{veg}} \cdot DIST_{VEG} + \beta^{dist_{hydro}} \cdot DIST_{HYDRO} \quad (\text{eq. 7})$$

<sup>4</sup> Loi de probabilité discrète utilisée généralement en écologie pour décrire le nombre d'observations d'un événement rare dans un intervalle défini (spatial ou temporel). Une loi de Poisson suit un paramètre  $\lambda$ , son intensité, qui correspond à la moyenne des observations et à leur variance.

#### Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

### 4.3.2 Résolution

L'estimation des paramètres  $\alpha$  et  $\beta$  de ce modèle a été réalisée dans un cadre Bayésien<sup>5</sup> 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  $\alpha$  et  $\beta$  du modèle suivent tous une loi normale non informative puisqu'aucune information a priori concernant l'impact des covariables sur l'occurrence des individus vivants et des carcasses n'a été intégrée.

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). Cette première méthode a été testée pour des versions préliminaires au modèle présenté dans ce rapport. Cependant si la méthode MCMC fournit une solution asymptotiquement exacte, elle est très 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. Après des essais préliminaires en MCMC qui se sont avérés non concluants d'un point de vue pratique (temps de calcul excessif), la résolution du modèle a été menée sous la méthode INLA. Pour ce faire le modèle a été implémenté 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 25km<sup>2</sup>. Les paramètres permettent de définir l'intensité en un 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 km<sup>2</sup> : multiplication de l'intensité du centroïde du pixel par l'aire d'un pixel – Intensité de cadavre dans un pixel de 24km<sup>2</sup> : multiplication du centroïde du pixel par la longueur de la portion de route traversant ce pixel).

### 4.3.3 Validation

L'ajustement du modèle a été évalué 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. 8). La précision de la prédiction sera évaluée par le RMSE, la racine de l'écart quadratique moyen entre les données simulées et les données observées, proche de 0 il indique une précision fine des données simulées par rapport aux données observées (eq. 9). Il indique la variabilité de la qualité de la prédiction et afin de faciliter son interprétation, il a également été exprimé comme un pourcentage de l'observation moyenne (eq. 10).

$$PBIAS = 100 \times \frac{\sum (pred - obs)}{\sum obs} \quad (eq.8)$$

$$RMSE = \sqrt{\frac{\sum ((pred - obs)^2)}{nobs}} \quad (eq.9)$$

$$RMSE . percent = \frac{RMSE}{\frac{\sum obs}{nobs}} \times 100 \quad (eq.10)$$

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).

<sup>5</sup> L'inférence Bayésienne est une méthode d'inférence statistique qui permet d'estimer des paramètres comme une distribution a posteriori calculée à partir des données ainsi que d'informations a priori (type de distribution que devrait prendre le paramètre, valeurs du paramètre déjà estimée par des experts). Ce type d'inférence donnera la courbe de probabilité a posteriori du paramètre sur la gamme de valeur que peut prendre celui-ci.

## 5 RÉSULTATS

### 5.1 Description des données

Entre 2014 et 2021, 17 107 chevreuils vivants ont été géoréférencés, soit en moyenne 2 138 chevreuils par an (Fig. 12). Pour les cadavres issus de collisions routière, en moyenne 482 cadavres ont été observés par année, ce qui représente 1.7 cadavres de chevreuil pour 10 km par an (Fig. 12).

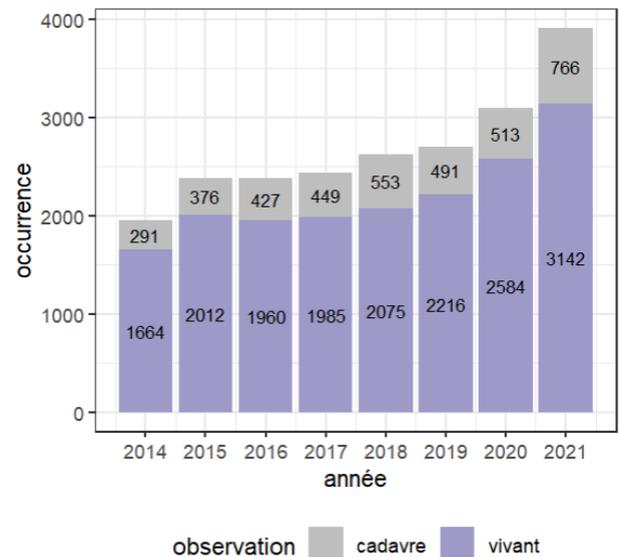


Figure 12 : Variation annuelle de l'occurrence des chevreuils vivants sur la région d'étude et de cadavre sur le réseau de la DIR Ouest

Les zones riches en observation de chevreuils vivants 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. 13.a). La carte est réalisée à une résolution de 24 km<sup>2</sup> et comporte quelques pixels qui regroupent une très grosse quantité de données (de 100 à 200 observations en 8 ans) (Fig. 13.a).

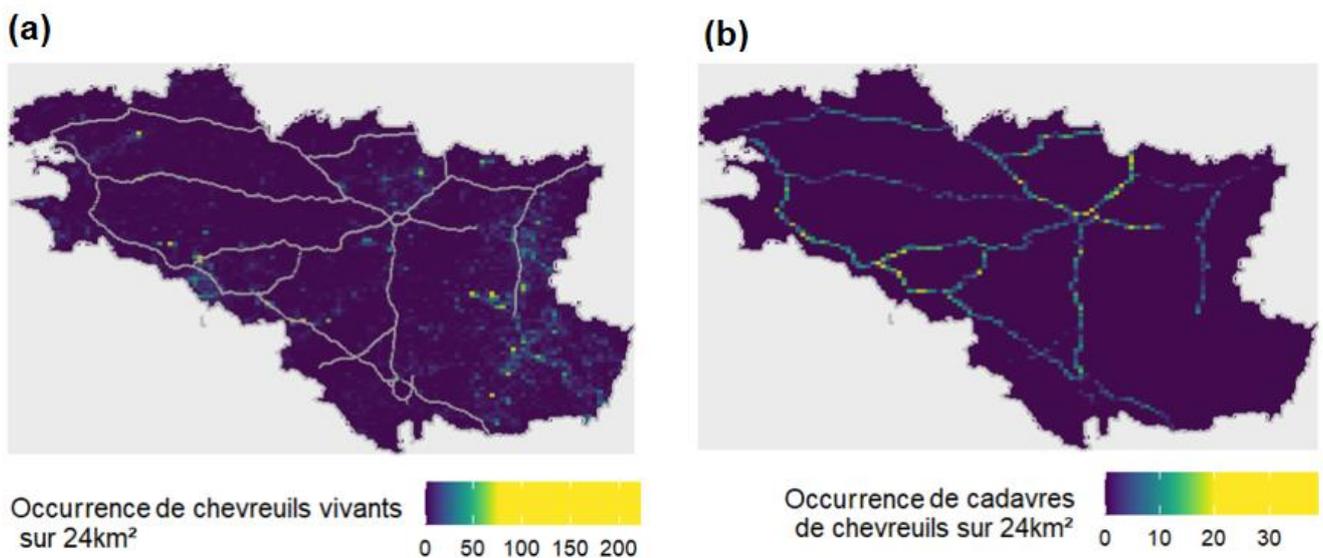


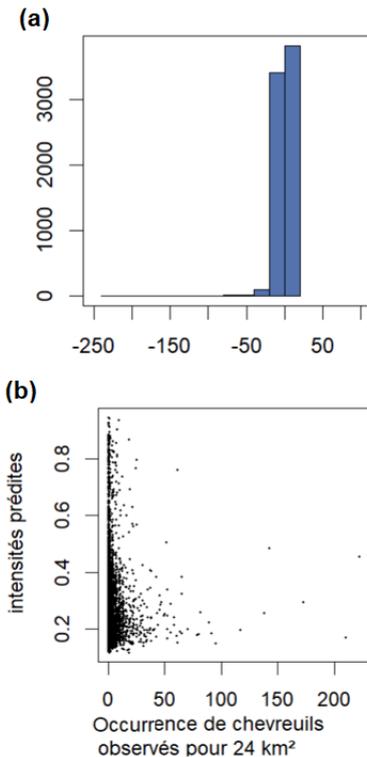
Figure 13 : Observations des chevreuils sur la période de 2014 à 2021, (a) vivants par le programme Vvisionature, (b) à l'état de cadavre sur le réseau de la DIR Ouest par les patrouilleurs

#### Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

Le nombre de cadavres de chevreuil collectés sur les routes varie entre 0 et 30 pour 24 km<sup>2</sup>. Les zones les plus mortifères se situent au Nord de Rennes, sur les routes partant de Lorient (Fig. 13.b).

## 5.2 Ajustement du modèle : pour les données Visionature de faune vivante



Le modèle fournit une abondance relative de chevreuils vivants observés par les naturalistes. Cette abondance relative ne peut pas être comparée à un nombre réel d'observations grâce aux indices du PBIAS et du RMSE, car l'unité de ce qu'on veut comparer n'est pas la même. La forme de la distribution des résidus (*donnée . prédite – donnée . observée*) ainsi que la forme de la relation entre l'abondance relative prédite (intensité prédite) et les observations peuvent cependant renseigner sur l'ajustement du modèle. La distribution des résidus est déséquilibrée, elle possède une grande queue décalée vers la gauche (Fig. 14.a), or dans le cas d'un modèle avec un bon ajustement à ses données, il serait attendu que la distribution des résidus soit équilibrée (de forme gaussienne). De plus, aucune tendance ni corrélation linéaire ne se dégage entre les intensités prédites et l'occurrence des observations de chevreuil vivant (Fig. 14.b).

Les données d'abondance relative prédites ne sont pas en adéquation avec les données observées. Cette conclusion est corroborée par l'inadéquation de la carte des données d'observations opportunistes de chevreuil prédites (Fig. 15) par rapport à la carte des observations opportunistes de chevreuil réelles (Fig. 13.a).

Figure 14 : Ajustement du modèle pour la prédiction de la distribution observée des chevreuils, (a) distribution des résidus, (b) Relation entre prédiction et observations

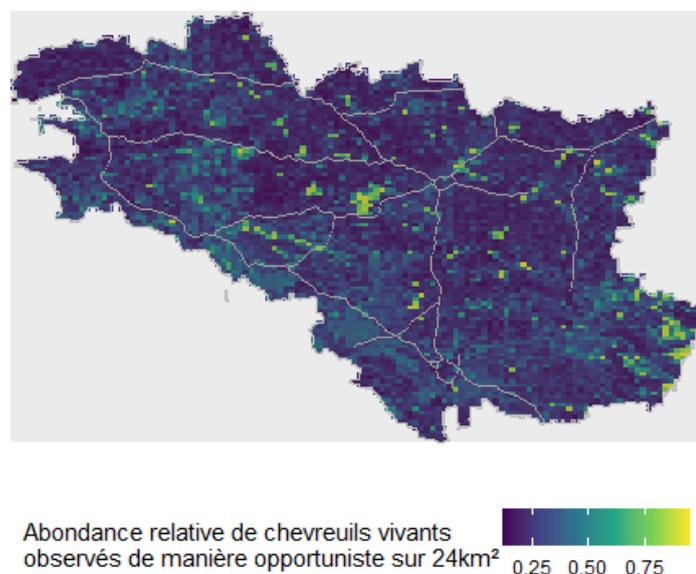
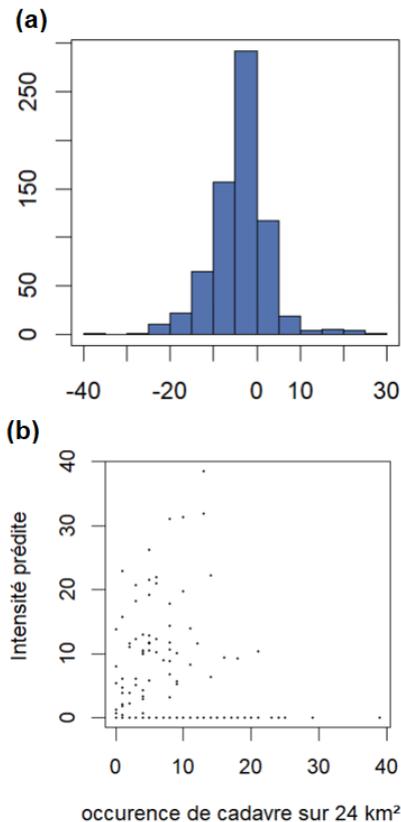


Figure 15 : Prédiction de l'abondance relative de points d'observation de chevreuil vivant par les bénévoles

### Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

## 5.3 Ajustement du modèle : pour les données DIRO de collision



L'intensité prédite des cadavres chevreuils géoréférencés ( ) 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 km<sup>2</sup>) (Fig. 19.c). Cependant les intensités de cadavre prédites sont plus clivées que les observations, on observe en effet une quasi-absence de valeurs d'intensité intermédiaires (Fig. 19.c et Fig 13.b). Les résidus ont une distribution équilibrée autour de la moyenne qui est légèrement inférieure à 0 (Fig. 16.a).

Excepté pour les valeurs d'intensité prédites comme nulles (ligne horizontale sur l'abscisse) alors qu'elles correspondent à des observations d'occurrences des cadavres de chevreuils allant de 0 à 40, une relation positive se dégage entre les intensités prédites et les observations de cadavres bien que la corrélation ne soit pas significative (Fig. 16.a).

Il y a donc un début d'ajustement du modèle entre les données de carcasses prédites et observées.

Figure 16 : 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

## 5.4 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. 17).

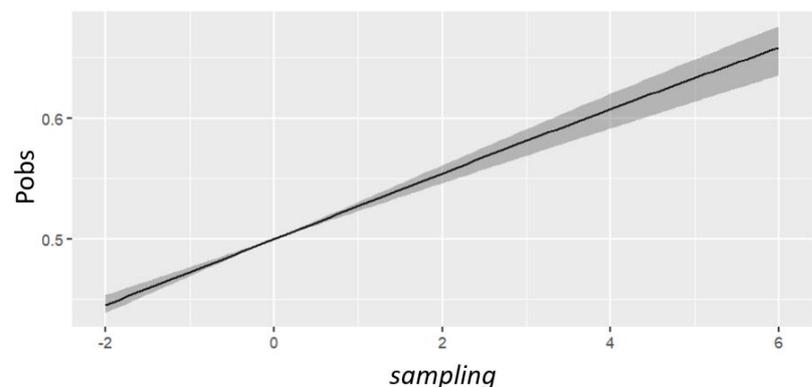


Figure 17: Relation entre la probabilité d'observer un chevreuil vivant et la covariable d'échantillonnage sampling

### Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

Pour les paramètres d'exposition, l'intercept  $\alpha$  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 de culture, de zones urbaines et une densité de haie élevées influent négativement sur l'abondance relative de chevreuils (Fig. 18). Une augmentation du recouvrement de prairie et de forêt implique en revanche une abondance relative de chevreuils plus élevée (Fig. 18).

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. 18). 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 (espaces végétaux naturels ou non, voir partie 4.2 'Prédicteurs du danger') et des cours d'eau sont plus dangereuses (Fig. 18).

Les prédicteurs biologiques de l'abondance de chevreuils vivants (exposition/ distribution) sont cohérents avec l'écologie du chevreuil. De plus les paramètres relatifs au danger routier sont également en adéquation avec les connaissances actuelles sur les caractéristiques dangereuses des routes.

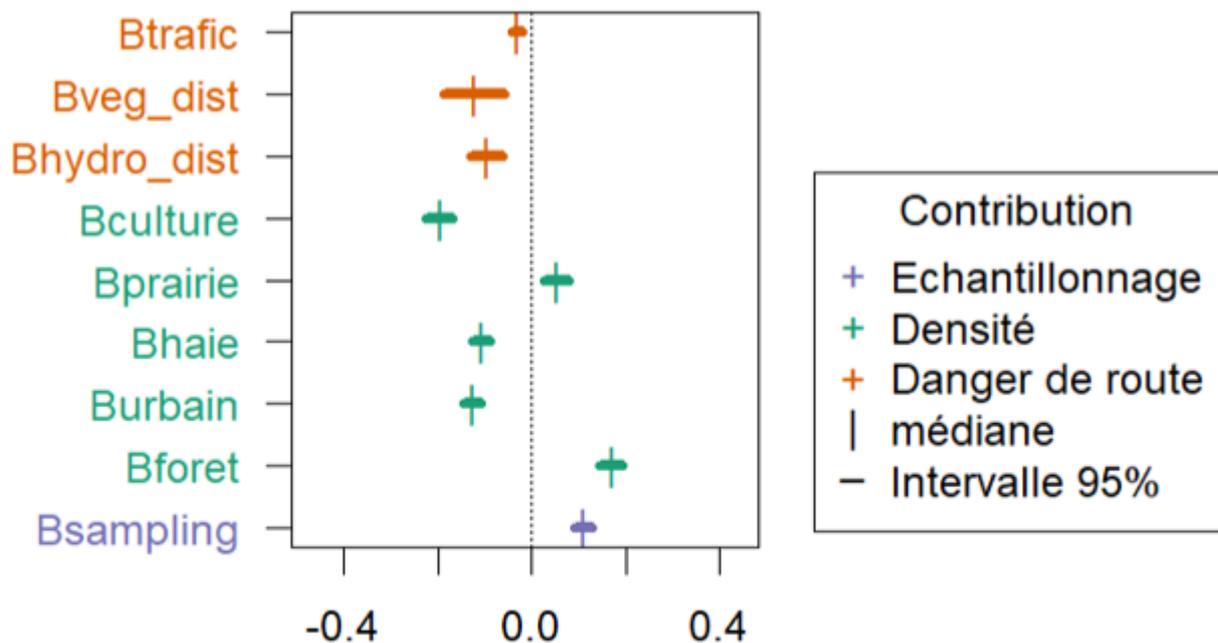


Figure 18 : Estimation des pentes des effet linéaires des covariables sur le danger de la route, la densité relative des chevreuils (exposition) et l'échantillonnage (les B correspondent à  $\beta$ )

### Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

## 5.5 Cartographie préliminaire du risque de collision routière

L'abondance relative des 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 forêts (Fig. 19.a et Fig. 1.a). Le danger intrinsèque à la route  $D_s$  est plus élevé sur les portions de routes supérieures à 90 km/h. Il est faible aux alentours des villes, notamment sur les périphériques qui comportent un fort trafic (Fig. 19.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, la route au 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. 19.c et Fig.13.b).

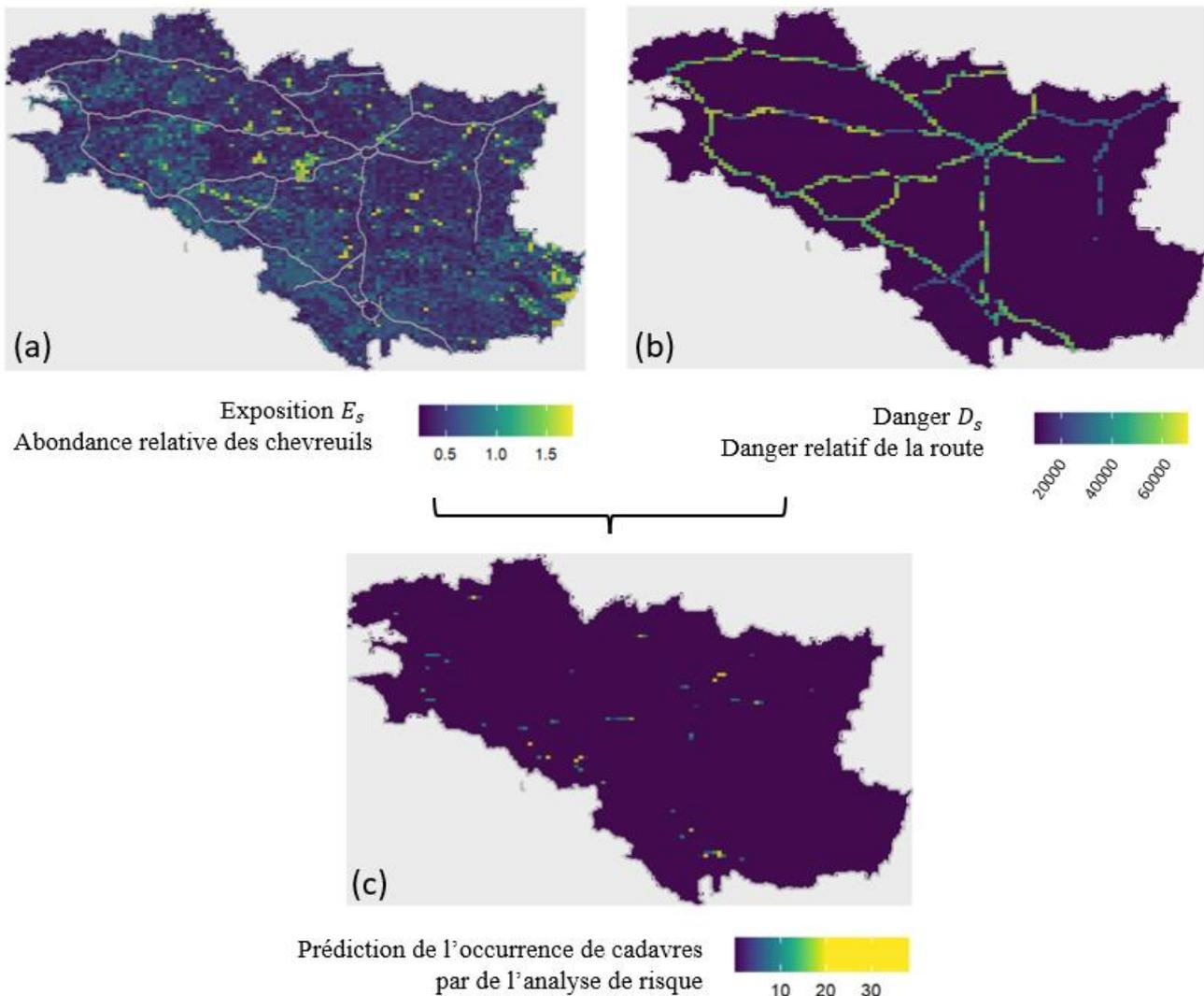


Figure 19 : Prédiction des collisions chevreuil-véhicule par l'analyse de risque entre 2014 et 2021 à résolution de 24km<sup>2</sup>, (a) Prédiction de l'exposition soit l'abondance relative de chevreuil déterminée par le modèle de niche, (b) Prédiction du danger intrinsèque à la route, (c) Prédiction de l'occurrence de cadavres de chevreuils

L'ajustement du modèle étant inadéquat, la qualité prédictive de celui-ci ne peut pas être satisfaisante. En effet pour la qualité prédictive temporelle, la prédiction des carcasses de chevreuils est sous-estimée (-80.92%) et son RMSE (0.87 en moyenne) comparé à la moyenne d'observation par maille de 24 km<sup>2</sup> représente 211.84% de la variabilité. Pour la qualité prédictive spatiale, la prédiction des carcasses est sous-estimée (-98.19%) et son RMSE (2.19 en moyenne) comparé à la moyenne d'observation par maille de 24 km<sup>2</sup> représente 156.41% de la variabilité.

### Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

## 6 BILAN

### 6.1 Discussion

#### 6.1.1 Résumé des résultats

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 dus à l'utilisation de données opportunistes et à la modélisation à grain fin à échelle régionale. **À ce stade du développement, l'ajustement du modèle et sa capacité prédictive sont insatisfaisants, mais ses paramètres sont biologiquement crédibles suggérant que l'architecture du modèle constitue une base de développement pertinente.** En effet, la couche d'exposition, filtrée par la couche d'effort d'échantillonnage, offre des prédictions qui s'ajustent difficilement aux observations opportunistes de faune vivante du programme Visionature. Cependant, l'ajustement de la couche représentant le danger (modélisation des cadavres) est plus cohérente avec les observations de carcasse de faune de la DIRO, bien qu'améliorable. **Par conséquent, la cartographie prédictive du risque de mortalité ambitionnée au lancement du projet exploratoire nécessite encore un travail de développement avant de pouvoir être proposée comme outil aux gestionnaires des routes et décideurs des politiques publiques d'aménagement du territoire.** Il convient cependant d'être prudent quant à l'opérationnalité d'un tel modèle. Ainsi on peut probablement espérer aboutir à un modèle plutôt bon à prédire dans l'espace à un gros grain spatial, mais la capacité prédictive dans le temps et le grain spatial fin semblent difficile à atteindre avec les données et outils de modélisation disponibles. Cela implique que ce modèle et ces données peuvent être utilisées dans un cadre de planification régionale, en identifiant les grandes zones accidentogènes. Il s'agira ensuite de réaliser un relevé plus fin et approfondi des collisions pour définir des localisations précises des aménagements. La cohérence biologique des paramètres suggère en revanche que la source des problèmes à résoudre ne réside ni dans une déficience des données, ni dans une sélection inadéquate des prédicteurs du modèle (paysager, danger routier). La discussion vise donc à détailler des déficiences actuelles de ce modèle exploratoire et présenter comment elles peuvent être dépassées.

#### 6.1.2 Questions autour des données opportunistes et de leur intégration

En terme d'ajustement aux données, le modèle souffre d'une difficulté d'étalonnage de l'effort d'échantillonnage. La prédiction des données opportunistes est trop peu impactée par l'effort d'échantillonnage. Elle correspond donc principalement à la couche d'exposition (d'abondance théorique de l'espèce étudiée). La stratégie d'intégration de l'effort d'échantillonnage a ici été d'intégrer les variables synthétiques de l'échantillonnage comme proposé dans l'étude de De Solan et al. (2019). Ces variables synthétiques permettent notamment d'identifier l'attractivité d'un site pour les observateurs bien qu'elles ne tiennent pas compte des variations individuelles de l'échantillonnage liées aux observateurs (connaissance naturalistes sur certains taxons spécifiques, technique d'observation, etc.). Cependant pour l'espèce ayant servi à tester notre modèle, le chevreuil, espèce commune et de grande taille, la variation individuelle de l'effort d'échantillonnage est peu susceptible de poser problème. Dans notre cas, la résolution (taille de la maille) des variables synthétiques de l'échantillonnage peut en revanche être une voie d'exploration à privilégier pour mieux intégrer la pression de l'échantillonnage d'observateurs opportunistes. En effet une maille trop grossière peut masquer des variations d'échantillonnage plus précises, cependant une maille trop fine n'aurait plus aucune qualité de synthèse, et, si elle permettrait de répliquer de manière quasi-exacte les données d'entrées, elle ne permettrait aucune généralisation du modèle (Botella et al., 2021). La maille utilisée dans notre modèle fait 24km<sup>2</sup>, et semble avoir une qualité de synthèse satisfaisante. Elle permet notamment d'illustrer que la pression d'échantillonnage est plus élevée dans les grandes villes et le long des axes routiers principaux ainsi que le long des côtes bretonnes qui constituent des espaces d'observations de la faune intéressants et fréquentés par les naturalistes. Le grain spatial des variables de synthèse de l'effort d'échantillonnage reste cependant déterminant et constitue un axe d'exploration de grand intérêt pour de futurs travaux.

#### **Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres**

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

### 6.1.3 L'apport possible de la spatialisation du modèle

Une autre approche pour améliorer l'ajustement du modèle est actuellement en cours d'étude par le réseau CiSStats. En effet, dans un modèle spatial, l'intégration de composante d'autocorrélation spatiale peut informer le réalisme des prédictions d'un modèle. Notamment car un champ d'autocorrélation spatiale peut capter les structures d'autocorrélation dues à des processus et des prédicteurs spatiaux non intégrés dans le modèle, soit à cause de l'indisponibilité de certaines données soit à cause de difficultés techniques de modélisation de certains processus. Pour citer quelques exemples, un champ d'autocorrélation spatiale pourrait capter la variation d'observation due à l'agrégation des animaux grégaires ou à des covariables très influentes sur le processus modélisé sur lesquelles les données sont indisponibles ou incomplètes (présence de clôture, obstacle sur la route, topologie de route, état végétal des dépendances). Ces exemples permettent uniquement d'illustrer biologiquement ce qui peut constituer des structures d'autocorrélation au sein d'un modèle. L'information captée par un champ d'autocorrélation spatiale ne permet pas de déterminer des liens de cause à effet entre les observations et des structures d'autocorrélation comme celles citées en exemple, mais cette information peut en revanche permettre d'améliorer les prédictions d'un modèle.

Il faut également noter que le modèle préliminaire actuel n'a été testé que sur l'espèce du chevreuil pour des raisons de simplicité. Une fois l'ajustement corrigé il sera nécessaire d'évaluer si la qualité d'ajustement et prédictive du modèle est conservée sur des espèces moins communes et pour lesquelles moins de données sont disponibles.

## 7 CONCLUSION

Le projet exploratoire COCPITT visait à étudier la faisabilité d'un modèle intégrant données opportunistes et données protocolées pour répondre à des questions de gestion et de politiques publiques concernant la mortalité de la faune par collision routière. Si le modèle ambitionné au début du projet n'est pas utilisable en l'état par les gestionnaires publics, son développement a permis d'ouvrir une réflexion sur la méthodologie pouvant être utilisée pour modéliser la mortalité routière de la faune à étendue régionale en tenant compte des données disponibles à cette étendue. La complexité d'une telle analyse de donnée réside notamment dans l'intégration de données opportunistes pour lesquelles le projet COCPITT propose un axe d'étude à partir de données de synthèse sur l'échantillonnage opportuniste. Ces données restent cependant difficiles à exploiter et l'amélioration de leur intégration dans un modèle nécessite une collaboration renforcée et sur la durée entre statisticiens, écologues et gestionnaires des infrastructures linéaires de transport. Le modèle a bénéficié d'une telle collaboration à travers le réseau de statisticiens CiSStats et du cadre offert par le programme Ittecop et son financement visant justement à mettre en relation ces différents acteurs de la sphère associative et de l'état. De nombreux échanges ont eu lieu entre statisticiens, référents des associations naturalistes, écologues afin de répondre aux enjeux de conservation et de gestion tout en valorisant les données opportunistes issues de sciences participatives. Cette collaboration a été fructueuse et a permis la construction d'un socle robuste de connaissance en vue d'un modèle complexe à approfondir sur le temps long. Le modèle développé au sein du projet COCPITT constitue une étape préliminaire, en effet les résultats présentés dans ce rapport offrent suffisamment de garantie de réussite pour conclure qu'une solution de modélisation peut permettre de répondre à l'objectif fixé avec des données non protocolées avec toutefois un coût de développement relativement élevé qu'il sera nécessaire de prendre en compte pour répondre à des demandes similaires ultérieurement.

## 8 ANNEXES

### Bibliographie

- 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., 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Guisan, A., Thuiller, W. & Zimmermann, N.E. (2017) *Habitat Suitability and Distribution Models: with Applications in R*, Cambridge University Press.
- 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.

#### **Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres**

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

- 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.
- Kelling, S., Johnston, A., Hochachka, W.M., Iloff, 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.
- Malo, J.E., Suárez, F. & Díez, A. (2004) Can we mitigate animal–vehicle accidents using predictive models? *Journal of Applied Ecology*, 41, 701–710.
- 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.
- 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.
- 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.
- 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-Aroita, 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., Carvalho, F. & Mira, A. (2011) How Long Do the Dead Survive on the Road? Carcass Persistence Probability and Implications for Road-Kill Monitoring Surveys. *PLOS ONE*, 6, e25383.
- 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.

### **Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres**

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

## Table des illustrations

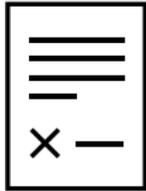
Figure 1: Site d'étude et son paysage : (a) Région d'acquisition des données, réseau routier de la DIR Ouest et cartographie du recouvrement forestier par km <sup>2</sup> (b) Recouvrement des zones urbaines par km <sup>2</sup> (c) Recouvrement des cultures par km <sup>2</sup> (d) Cartographie de l'effort d'échantillonnage par maille 5x5km (e) Recouvrement des prairies par km <sup>2</sup> (f) Densité de haie par km <sup>2</sup> .....	11
Figure 2: Composition des données de carcasse d'espèces sauvages observées par les agents de la DIR Ouest de 2014 à 2021 (à l'exception des amphibiens et des reptiles). Le symbole ° regroupe les espèces dont l'observation est anecdotique (<60 observations) .....	13
Figure 3: Occurrence de carcasses sur le réseau routier de la DIRO de 2014 à 2021 .....	14
Figure 4 : Nombre d'observations d'animal vivant à une localisation précise de la base de données Visionature de 2014 à 2021 pour les Pays de la Loire et la Bretagne .....	16
Figure 5 : Schéma du modèle théorique exhaustif de la mortalité de la faune par collision routière et frise synthétique des dimensions explorées dans le modèle du projet COCPITT.....	17
Figure 6 : Groupes d'espèces pouvant comporter des facteurs de risque de collision similaires, déterminés par classification hiérarchique ascendante .....	19
Figure 7: Schéma de principe d'une analyse de risque couplant l'exposition (a), placement d'un animal de manière qui l'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) .....	20
Figure 8 : Schéma des données d'ajustement du modèle et frise synthétique des données disponibles conservées dans la modèle du projet COCPITT .....	21
Figure 9 : Tableau des corrélations entre les variables de synthèse sur la base de données Visionature .....	23
Figure 10 : Cartographie des scores du premier axe PC1 de l'ACP (les valeurs négatives correspondent à une forte pression d'échantillonnage) .....	23
Figure 11 : Schéma du modèle réalisé (à gauche) et correspondance avec l'écriture mathématique du modèle (à droite) .....	26
Figure 12 : Variation annuelle de l'occurrence des chevreuils vivants sur la région d'étude et de cadavre sur le réseau de la DIR Ouest.....	29
Figure 13 : Observations des chevreuils sur la période de 2014 à 2021, (a) vivants par le programme Visionature, (b) à l'état de cadavre sur le réseau de la DIR Ouest par les patrouilleurs .....	29
Figure 14 : Ajustement du modèle pour la prédiction de la distribution observée des chevreuils, (a) distribution des résidus, (b) Relation entre prédiction et observations.....	30
Figure 15 : Prédiction de l'abondance relative de points d'observation de chevreuil vivant par les bénévoles .....	30
Figure 16 : 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 .....	31
Figure 17: Relation entre la probabilité d'observer un chevreuil vivant et la covariable d'échantillonnage sampling .....	31
Figure 18 : 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 $\beta$ ) .....	32
Figure 19 : Prédiction des collisions chevreuil-véhicule par l'analyse de risque entre 2014 et 2021 à résolution de 24km <sup>2</sup> , (a) Prédiction de l'exposition soit l'abondance relative de chevreuil déterminée par le modèle de niche, (b) Prédiction du danger intrinsèque à la route, (c) Prédiction de l'occurrence de cadavres de chevreuils .....	33

### Collisions Opportunistes et Collisions Protocoles liées aux Infrastructures de Transport Terrestres

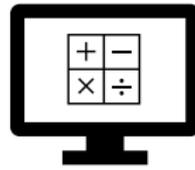
Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

# Glossaire

## Ecrire modèle



## Calcul des paramètres les plus vraisemblables



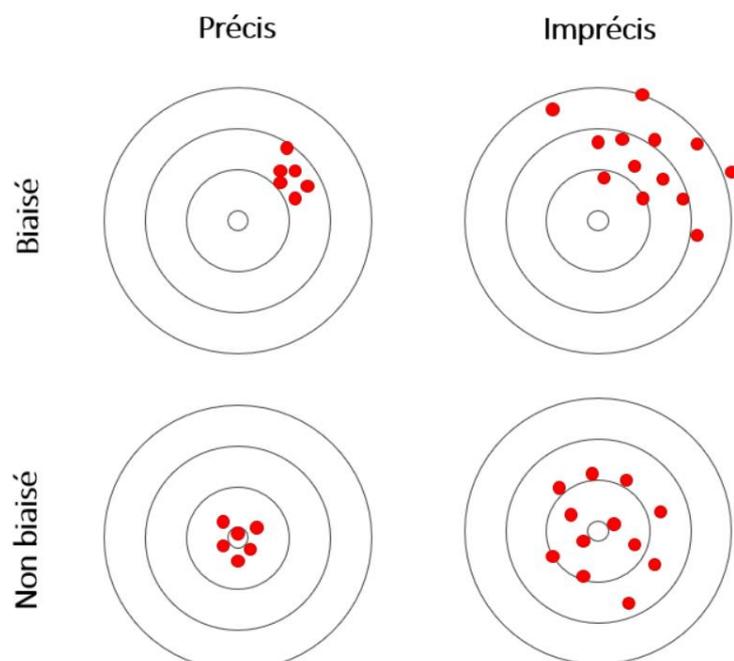
## Résultats - prédictions

- Cartes prédictives
- Impacts des variables explicatives
- Intervalle de confiance des variables explicatives

On met des règles (mathématiques) sur les relations entre les variables explicatives et données.

Une méthode d'approximation

- **Covariable** : variable explicative ;
- **Intercept** : ordonnée à l'origine / estimation constante la plus juste s'il n'y avait aucune covariable ;
- **Coefficient beta**: pente d'une régression linéaire ;
- **Utilisation du log** : permet de régler un paramètre pour qu'il soit forcément supérieur à 0 ;
- **Utilisation du logit()** : la fonction logit donne un résultat entre 0 et 1 ;
- **Résidus** : variance non expliquée par le modèle/ erreur du modèle / différence entre les observations et les données répliquées ;
- **Biais et précision** :

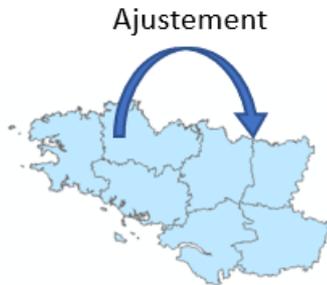


## **Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres**

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

## Ajustement et qualité prédictive :

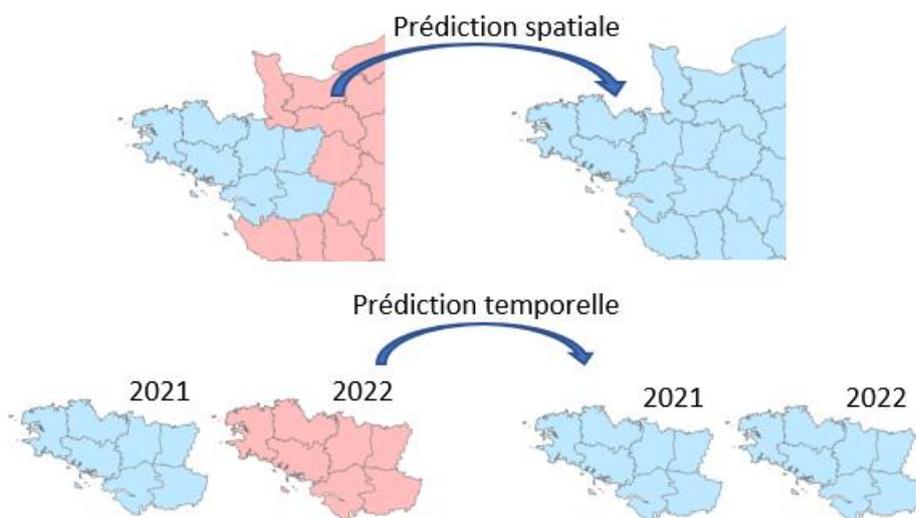
L'**ajustement** d'un modèle correspond à sa qualité d'ajustement aux données ayant servi à l'estimer. Il peut être mesuré en évaluant la qualité de la capacité à reprédire les données utilisées lors de son estimation.



Est-on capable de reprédire les données utilisées lors de l'estimation du modèle ?

*Sur l'illustration : peut on reproduire les données sur la région et la période informées de données (en bleu)*

La **qualité prédictive** d'un modèle correspond à sa capacité à prédire le processus modélisé (ici le risque de collision routière) sur une période et/ou un espace pour lesquels aucune donnée n'est disponible (c'est à dire dont aucune donnée n'a été utilisée pour estimer les paramètres du modèle).



Quelle est la qualité des prédictions si on les étend à des régions et/ou des périodes sur lesquelles le modèle n'a pas été ajusté ?

*Sur l'illustration : Prédiction spatiale : Quelle est la qualité des prédictions de risque de collision routière sur les régions pour lesquelles on a pas d'information (départements en rouge sur les cartes de gauche) ?*

*Prédiction temporelle : Quelle est la qualité des prédictions de risque de collision routière pour les années pour lesquelles on n'a pas d'information (départements en rouge sur les cartes de gauche) ?*

En règle générale, pour pouvoir évaluer la qualité prédictive d'un modèle, on choisit de l'ajuster avec moins de données que celles disponibles (par exemple on retire des territoires et/ou des années aux données d'ajustement) puis on évalue la qualité prédictive du modèle sur les données retirées au préalable. Les données d'ajustement sont appelées données de training et celle de validation de qualité prédictive données de test.

**A noter :** Un modèle peut avoir un très bon ajustement et une mauvaise qualité prédictive. Cependant si l'ajustement d'un modèle est mauvais, sa qualité prédictive le sera aussi.

### Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

# Classification des espèces

## Tableaux des traits

Espece	Taille	Poids	regime alimentaire	gregarisme	foraging	deplacement	nocturnalité
Lievre		55	4,75	1	2 glanage	rapide	Y
Lapin		41	1,7	1	1 glanage	lent	Y
Renard		74	7,5	8	2 chasse_active	rapide	Y
Blaireau		73	10	7	1 opportuniste	rapide	Y
Chevreuril		95	22	1	1 broutage	rapide	Y
Sanglier		135	150	3	1 opportuniste	rapide	Y
Cerf		200	162	1	1 broutage	rapide	Y
Fouine		47	1,7	8	3 chasse_active	rapide	Y
Martre		56,2	1,3	10	3 chasse_active	rapide	Y
Belette		22	0,215	10	3 chasse_active	rapide	Y
Ecureuil		22,5	0,275	3	3 opportuniste	rapide	N
Herisson		24,5	1,1	7	3 opportuniste	lent	Y
Ragondin		52,5	6,5	2	1 glanage	lent	N
Putois		50	1,2	8	3 chasse_active	rapide	Y
Rat musqué		55,5	1,3	1	2 glanage	lent	Y
Rat		21,5	0,44	5	1 opportuniste	lent	Y
Hermine		23,5	0,26	10	3 chasse_active	rapide	Y

espece	taille	poids	regime alimentaire	gregarisme	foraging	type_envol	nocturnalite
Chouette sp		37	381,666667	10	3 affut_maraude	plane	Y
Pigeon		40,5	354,5	1	1 terre	battu	N
Pigeon ramier		40,5	487	1	1 terre	battu	N
Buse variable		54	893,5	10	3 affut_maraude	plane	N
Merle noir		26,25	95	6	3 arbres_buissons	battu	N
Effraie des clochers		40	321	10	3 affut_maraude	plane	Y
Corvidé		46	499	5	1 terre	battu	N
Mouette rieuse		43	260	6	1 terre	battu	N
Pie bavarde		45,5	200,5	6	2 terre	plane	N
Goéland argenté		57	821	7	2 terre	battu	N
Epervier d'Europe		35	227,5	10	3 affut_maraude	plane	N
Canard colvert		55	1335	3	2 terre	battu	N
Chouette hulotte		34	496,5	9	3 affut_maraude	plane	Y
Grive musicienne		21	78,5	6	3 arbres_buissons	battu	N
Hibou moyenduc		34	327,5	10	3 affut_maraude	plane	Y
Geai des chênes		33,5	165	5	2 arbres_buissons	battu	N
Perdrix grise		30	455	3	1 terre	battu	N
Faisan de Colchide		72,5	1750	1	2 terre	battu	N

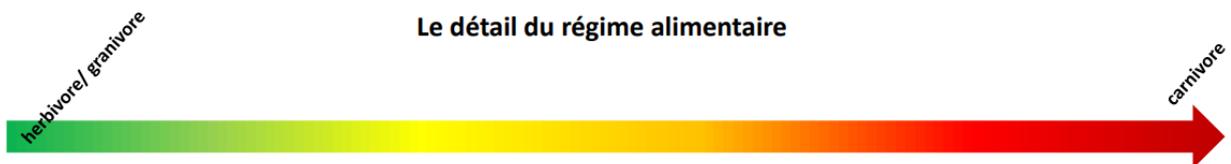
### Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

## Description des traits

### Résumé des traits choisis :

- **Taille**  
quantitatif
- **Poids**  
quantitatif
- **Régime**  
quantitatif (voir diapositive suivante)
- **Grégarisme**  
quantitatif (1 grégaire, 2 moyennement grégaire, 3 non grégaire)
- **Foraging (recherche de nourriture)**  
qualitatif : Mammifère. broutage - glanage - opportuniste - chasse\_active  
Avifaune. terre - arbre\_buisson - affut\_maraude
- **Type\_envol (avifaune) / déplacement (mammifère)**  
qualitatif : Mammifère. lent - rapide  
Avifaune. battu - plané
- **Nocturnalité**  
qualitatif (Y oui, N non)



#### Avifaune

	1	2	3	4	5	6	7	8	9	10
Faisan de Colchide		Canard colvert		Geai des chênes	Mouette rieuse	Goéland argenté		Chouette hulotte	Effraie des clochers	
Pigeon		Perdrix grise		Corvidé	Pie bavarde				Epervier d'Europe	
Pigeon ramier					Merle noir				Hibou moyenduc	
					Grive musicienne				Chouette sp	
									Buse variable	

#### Mammifères

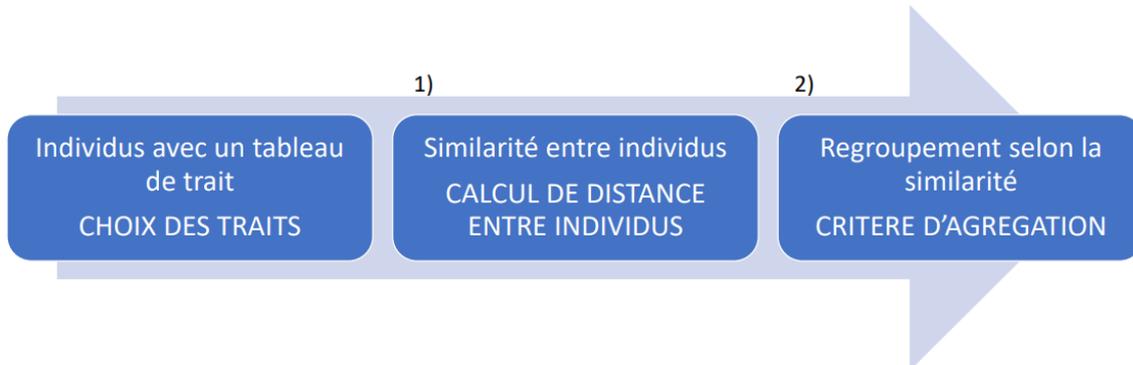
	1	2	3	4	5	6	7	8	9	10
lievre	ragondin	sanglier		rat		hérisson	putois			martre
lapin		écureuil				blaireau	fouine			belette
chevreuil							renard			hermine
cerf										
rat musqué										

## Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

# Démarche de classification hiérarchique ascendante

La démarche pour réaliser des groupes à partir de leur similarité sur des traits donnés



1) Comment on calcule la distance (1- similarité) entre les individus ?

2) Avec ces distances comment on regroupe les individus ?

## 1) Comment on calcule la distance/similarité entre les individus?

Dans notre cas on a des variables qualitatives et quantitatives, un indice de similarité qui correspond à ce type de données est l'indice de Gower.

### Distance de Gower

En 1971, Gower a proposé un indice de similarité qui porte son nom <sup>(1)</sup>. L'objectif de cet indice consiste à mesurer dans quelle mesure deux individus sont semblables. L'indice de Gower varie entre 0 et 1. Si l'indice vaut 1, les deux individus sont identiques. À l'opposé, s'il vaut 0, les deux individus considérés n'ont pas de point commun. Si l'on note  $S_g$  l'indice de similarité de Gower, la distance de Gower  $D_g$  s'obtient simplement de la manière suivante :  $D_g = 1 - S_g$ . Ainsi, la distance sera nulle entre deux individus identiques et elle sera égale à 1 entre deux individus totalement différents. Cette distance s'obtient sous R avec la fonction `daisy` du package `cluster`.

L'indice de similarité de Gower entre deux individus  $x_1$  et  $x_2$  se calcule de la manière suivante :

$$S_g(x_1, x_2) = \frac{1}{p} \sum_{j=1}^p s_{12j}$$

$p$  représente le nombre total de caractères (ou de variables) descriptifs utilisés pour comparer les deux individus <sup>(2)</sup>.  $s_{12j}$  représente la similarité partielle entre les individus 1 et 2 concernant le descripteur  $j$ . Cette similarité partielle se calcule différemment s'il s'agit d'une variable qualitative ou quantitative :

- **variable qualitative** :  $s_{12j}$  vaut 1 si la variable  $j$  prend la même valeur pour les individus 1 et 2, et vaut 0 sinon. Par exemple, si 1 et 2 sont tous les deux « grand », alors  $s_{12j}$  vaudra 1. Si 1 est « grand » et 2 « petit »,  $s_{12j}$  vaudra 0.
- **variable quantitative** : la différence absolue entre les valeurs des deux variables est tout d'abord calculée, soit  $|y_{1j} - y_{2j}|$ . Puis l'écart maximum observé sur l'ensemble du fichier est déterminé et noté  $R_j$ . Dès lors, la similarité partielle vaut  $s_{12j} = 1 - |y_{1j} - y_{2j}| / R_j$ .

Distance de Gower



Similarité de Gower



### Exemple de calcul

arbre	Taille	À fleur	milieu	ligneuse		
A	10	O	humide	N		
B	25	O	sec	N		
C	500	N	sec	O		

	s taille	s_a fleur	s milieu	s ligneuse	Similarité	Distance
A - B	0.96938776	1	0	1	0.74234694	0.25765306
A - C	0	0	0	0	0	1
B - C	0.03061224	0	1	0	0.25765306	0.74234694

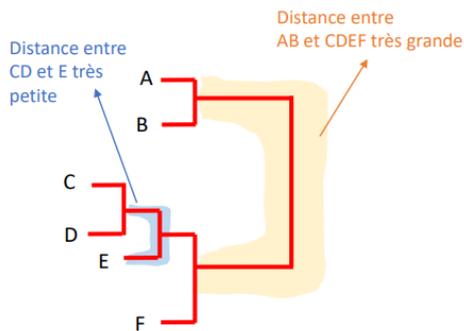
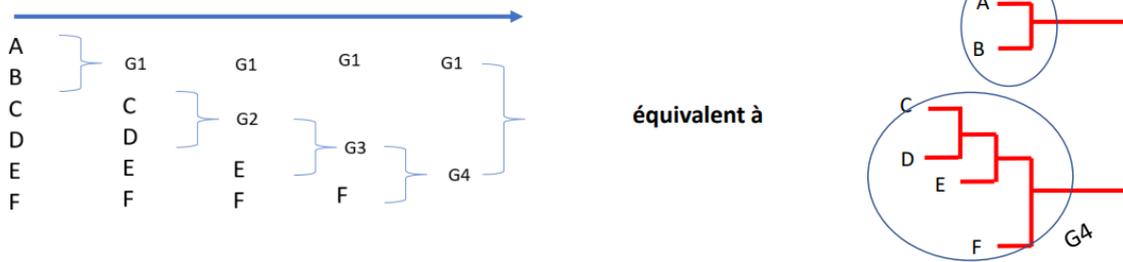
Source : <https://larmarange.github.io/analyse-R/classification-ascendante-hierarchique.html>

## Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

## 2) Avec ces distances comment on regroupe les individus ?

On agrège les individus dans des groupes progressivement (classification hiérarchique ascendante) :  
 À chaque itération on regroupe les 2 individus ou groupes les plus similaires :



Les distance des traits en rouges correspondent aux distances entre groupes

À vue d'œil on a 2 groupes très différents :

- A et B
- C, D, E et F

## Scripts R du modèle

### ### librairies et données générales ----

```
library(sf) ; library(inlabru) ; library(INLA) ; library(ggplot2) ; library(maptools) ; library(igraph)
library(dplyr) ; library(raster) ; library(tidygraph) ; library(sfnetworks) ; library(tidyverse) ; library(ade4) ;
library(fields)
```

```
vivant_30sp=read.csv("input/input_30sp/DFvisionature_format_final.csv", h=T)
```

```
collision_30sp=read.csv("input/input_30sp/roadkill_format_final.csv", h=T)
```

```
id_sp=read.csv("input/input_30sp/ID_NAME_sp_grp.csv", h=T)
```

### ###covariables mise au format inlabru ----

```
foretR <- raster("input/input_30sp/foret_landcover_1km.tif")
```

```
foretSG <- as(scale(foretR), 'SpatialGridDataFrame')
```

```
prairieR <- raster("input/input_30sp/prairie_landcover_1km.tif")
```

```
prairieSG <- as(scale(prairieR), 'SpatialGridDataFrame')
```

```
traficR <- raster("input/input_30sp/trafic_1km_meshRoad.tif")
```

```
traficSG <- as(traficR/5000, 'SpatialGridDataFrame') #on ne scale pas le trafic car il corrrespond à qqc
de déterministe on travaille cependant par tranche de 5000 voitures
```

```
samplingR <- raster("input/input_30sp/sampling_axis1_1km.tif")
```

```
samplingR=-1*samplingR ### on met sampling dans le bon sens (plus la valeur est haute plus l'échant
est fort)
```

```
samplingSG <- as(scale(sqrt(samplingR-summary(samplingR)[1,])), 'SpatialGridDataFrame')
```

```
cultureR <- raster("input/input_30sp/culture_landcover_1km.tif")
```

```
cultureSG <- as(scale(cultureR), 'SpatialGridDataFrame')
```

```
haieR <- raster("input/input_30sp/densite_haie_1km.tif")
```

```
haieSG <- as(scale(haieR), 'SpatialGridDataFrame')
```

```
hydro_distR <- raster("input/input_30sp/raster_hydro_dist_50m.tif")
```

```
hydro_distSG <- as(scale(hydro_distR), 'SpatialGridDataFrame')
```

```
haie_distR <- raster("input/input_30sp/raster_veg_dist_1km.tif")
```

```
haie_distSG <- as(scale(haie_distR), 'SpatialGridDataFrame')
```

```
urbainR <- raster("input/input_30sp/urbain_landcover_1km.tif")
```

```
urbainSG <- as(scale(urbainR), 'SpatialGridDataFrame')
```

```
vitesseR <- raster("input/input_30sp/vitesse_quali_1km_meshRoad.tif")
```

```
vitesseSG <- as(vitesseR, 'SpatialGridDataFrame')
```

```
vitesse90SG<-vitesseSG
```

```
vitesse90SG@data$vitesse_quali_1km_meshRoad<-as.vector(ifelse(vitesse90SG@data==1 &
lis.na(vitesse90SG@data),1,0))
```

### Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

```
vitesse100SG<-vitesseSG
vitesse100SG@data$vitesse_quali_1km_meshRoad<-as.vector(ifelse(vitesse100SG@data==3 &
lis.na(vitesse100SG@data),1,0))
vitesse120SG<-vitesseSG
vitesse120SG@data$vitesse_quali_1km_meshRoad<-as.vector(ifelse(vitesse120SG@data==2 &
lis.na(vitesse120SG@data),1,0))
vitesse100SG@data$vitesse_quali_1km_meshRoad=vitesse100SG@data$vitesse_quali_1km_mesh
Road+vitesse120SG@data$vitesse_quali_1km_meshRoad
plot(vitesse100SG)
```

```
f.vitesse90 <- function(x, y) {
  spp <- SpatialPoints(data.frame(x = x, y = y), proj4string = fm_sp_get_crs(vitesse90SG))
  proj4string(spp) <- fm_sp_get_crs(vitesse90SG)
  v <- over(spp, vitesse90SG)
  if (any(is.na(v$vitesse_quali_1km_meshRoad))) {
    v$vitesse_quali_1km_meshRoad[is.na(v$vitesse_quali_1km_meshRoad)] <- 0
  }
  return(v$vitesse_quali_1km_meshRoad)
}
```

```
f.vitesse100 <- function(x, y) {
  spp <- SpatialPoints(data.frame(x = x, y = y), proj4string = fm_sp_get_crs(vitesse100SG))
  proj4string(spp) <- fm_sp_get_crs(vitesse100SG)
  v <- over(spp, vitesse100SG)
  if (any(is.na(v$vitesse_quali_1km_meshRoad))) {
    v$vitesse_quali_1km_meshRoad[is.na(v$vitesse_quali_1km_meshRoad)] <- 0
  }
  return(v$vitesse_quali_1km_meshRoad)
}
```

```
f.vitesse120 <- function(x, y) {
  spp <- SpatialPoints(data.frame(x = x, y = y), proj4string = fm_sp_get_crs(vitesse120SG))
  proj4string(spp) <- fm_sp_get_crs(vitesse120SG)
  v <- over(spp, vitesse120SG)
  if (any(is.na(v$vitesse_quali_1km_meshRoad))) {
    v$vitesse_quali_1km_meshRoad[is.na(v$vitesse_quali_1km_meshRoad)] <- 0
  }
  return(v$vitesse_quali_1km_meshRoad)
}
```

```
f.foret <- function(x, y) {
  spp <- SpatialPoints(data.frame(x = x, y = y), proj4string = fm_sp_get_crs(foretSG))
```

### **Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres**

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

```

proj4string(spp) <- fm_sp_get_crs(foretSG)
v <- over(spp, foretSG)
if (any(is.na(v$foret_landcover_1km))) {
  v$foret_landcover_1km[is.na(v$foret_landcover_1km)] <- 0# <- inlabru::bru_fill_missing(foretSG,
spp, v$foret_landcover_1km)
}
return(v$foret_landcover_1km)
}

f.prairie <- function(x, y) {
  spp <- SpatialPoints(data.frame(x = x, y = y), proj4string = fm_sp_get_crs(prairieSG))
  proj4string(spp) <- fm_sp_get_crs(prairieSG)
  v <- over(spp, prairieSG)
  if (any(is.na(v$prairie_landcover_1km))) {
    v$prairie_landcover_1km[is.na(v$prairie_landcover_1km)] <- 0
  }
  return(v$prairie_landcover_1km)
}

f.trafic <- function(x, y) {
  spp <- SpatialPoints(data.frame(x = x, y = y), proj4string = fm_sp_get_crs(traficSG))
  proj4string(spp) <- fm_sp_get_crs(traficSG)
  v <- over(spp, traficSG)
  if (any(is.na(v$trafic_1km_meshRoad))) {
    v$trafic_1km_meshRoad[is.na(v$trafic_1km_meshRoad)] <- 0
  }
  return(v$trafic_1km_meshRoad)
}

f.sampling <- function(x, y) {
  spp <- SpatialPoints(data.frame(x = x, y = y), proj4string = fm_sp_get_crs(samplingSG))
  proj4string(spp) <- fm_sp_get_crs(samplingSG)
  v <- over(spp, samplingSG)
  if (any(is.na(v$layer))) {
    v$layer[is.na(v$layer)] <- 0
  }
  return(v$layer)
}

f.culture <- function(x, y) {
  spp <- SpatialPoints(data.frame(x = x, y = y), proj4string = fm_sp_get_crs(cultureSG))
  proj4string(spp) <- fm_sp_get_crs(cultureSG)

```

### **Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres**

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

```

v <- over(spp, cultureSG)
if (any(is.na(v$culture_landcover_1km))) {
  v$culture_landcover_1km[is.na(v$culture_landcover_1km)] <- 0
}
return(v$culture_landcover_1km)
}

f.haie <- function(x, y) {
  spp <- SpatialPoints(data.frame(x = x, y = y), proj4string = fm_sp_get_crs(haieSG))
  proj4string(spp) <- fm_sp_get_crs(haieSG)
  v <- over(spp, haieSG)
  if (any(is.na(v$densite_haie_1km))) {
    v$densite_haie_1km[is.na(v$densite_haie_1km)] <- 0
  }
  return(v$densite_haie_1km)
}

f.hydro_dist <- function(x, y) {
  spp <- SpatialPoints(data.frame(x = x, y = y), proj4string = fm_sp_get_crs(hydro_distSG))
  proj4string(spp) <- fm_sp_get_crs(hydro_distSG)
  v <- over(spp, hydro_distSG)
  if (any(is.na(v$raster_hydro_dist_50m))) {
    v$raster_hydro_dist_50m[is.na(v$raster_hydro_dist_50m)] <- 0
  }
  return(v$raster_hydro_dist_50m)
}

f.haie_dist <- function(x, y) {
  spp <- SpatialPoints(data.frame(x = x, y = y), proj4string = fm_sp_get_crs(haie_distSG))
  proj4string(spp) <- fm_sp_get_crs(haie_distSG)
  v <- over(spp, haie_distSG)
  if (any(is.na(v$raster_veg_dist_1km))) {
    v$raster_veg_dist_1km[is.na(v$raster_veg_dist_1km)] <- 0
  }
  return(v$raster_veg_dist_1km)
}

f.urbain <- function(x, y) {
  spp <- SpatialPoints(data.frame(x = x, y = y), proj4string = fm_sp_get_crs(urbainSG))
  proj4string(spp) <- fm_sp_get_crs(urbainSG)
  v <- over(spp, urbainSG)
  if (any(is.na(v$urbain_landcover_1km))) {

```

### **Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres**

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

```

v$urbain_landcover_1km[is.na(v$urbain_landcover_1km)]
inlabru:::bru_fill_missing(urbainSG, spp, v$urbain_landcover_1km)
}
return(v$urbain_landcover_1km)
}

```

&lt;-

0#

&lt;-

### ### domaine-----

```

domaine = st_read("input/input_30sp/emprise_maillage_1km_g2.shp")
routes = st_read("input/input_30sp/diro_simpl.shp")

```

```

coordDOM <- st_coordinates(domaine)
bndint = inla.nonconvex.hull(as.matrix(coordDOM[,1:2]), convex=-.11)
size_mesh=5000
mesh = inla.mesh.2d(boundary = bndint,max.edge =size_mesh)
# par(mar = rep(0.5, 4))
# plot(mesh, main = "", asp = 1)
# plot(domaine, add=T)

```

```

st_crs(routes)=NA
routesSL <- as(routes,"Spatial")
st_crs(domaine) <- NA
domaineSP <- as_Spatial(domaine)

```

### #### données 1sp ####

```

currentID_sp=21 # identifiant du chevreuil
currentNAME_sp=id_sp$nom_sci[id_sp$ID_spe==currentID_sp]
vivant=vivant_30sp[vivant_30sp$ID_sp==currentID_sp,]
collision=collision_30sp[collision_30sp$ID_sp==currentID_sp,]
vivant$SAMPLING=f.sampling(vivant$x,vivant$y)
collision$trafic=f.trafic(collision$x,collision$y) #inutile
collision$hydro_dist=f.hydro_dist(collision$x,collision$y) #inutile
collision$haie_dist=f.haie_dist(collision$x,collision$y) #inutile
collision$vitesse90=f.vitesse90(collision$x,collision$y) #inutile
collision$vitesse100=f.vitesse100(collision$x,collision$y) #inutile

```

```

datcoll <- SpatialPointsDataFrame(collision[,c("x", "y")], collision[,c(1:3,5:7,10,12,13:ncol(collision))])
datviv <- SpatialPointsDataFrame(vivant[,c("x", "y")], vivant[,c(1:3,5:7,10,13,14:ncol(vivant))])

```

### Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

```
lignes <- routesSL
poly <- domaineSP
```

```
##### Run #####
```

```
inla.set.control.inla.default()
```

```
pobs = function(SAMPLING, Bsampling){
  (1/(1+exp(-(abs(Bsampling)*SAMPLING))))
}
```

```
## exemple d'implementation de champs d autocorrelation spatiale
```

```
# matern <- inla.spde2.pcmatern(mesh,
# prior.sigma = c(0.1, 0.99),
# prior.range = c(1000, 0.01)) #c(100000, 0.99) c(1000, 0.01)
```

```
cmp <- ~ -1 +ALPHA(1)+
  ALPHA.vitesse90(main = f.vitesse90(x,y))+
  ALPHA.vitesse100(main = f.vitesse100(x,y))+
  Bforet(main = f.foret(x,y), model = "linear")+
  Burbain(main = f.urbain(x,y), model = "linear")+
  Bhaie(main = f.haie(x,y), model = "linear")+
  Bprairie(main = f.prairie(x,y), model = "linear")+
  Bculture(main = f.culture(x,y), model = "linear")+
  Bsampling(1)+
  Bhydro_dist(main = f.hydro_dist(x,y), model = "linear")+
  Bveg_dist(main = f.haie_dist(x,y), model = "linear")+
  Btrafic(main = f.trafic(x,y), model="linear")#+
#mySPDE(main=coordinates, model=matern)
```

```
#### EQUATIONS
```

```
## formule de la modélisation des vivants
```

```
form1= coordinates + SAMPLING ~
  -1+ALPHA+
  Bhaie+Bforet+Burbain+Bculture+Bprairie+
  log(pobs(SAMPLING, Bsampling)) #+mySPDE
```

```
## domaine sur lequel modéliser les vivants
```

### **Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres**

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude

```
dom1=list(coordinates = mesh,
          SAMPLING = INLA::inla.mesh.1d(seq(-2.5,6,length.out=20))) #seq(-1.5,10,length.out=20) seq(-
2.5,6,length.out=20)
```

#### ## formule de la modélisation des carcasses

```
form2= coordinates ~
-1 +ALPHA+ALPHA.vitesse90+ALPHA.vitesse100+
Bhaie+Bforet+Burbain+Bculture+Bprairie+
Btrafic + Bhydro_dist + Bveg_dist #+mySPDE
```

#### ## domaine sur lequel modéliser les carcasses

```
dom2=list(coordinates = mesh)
```

```
lik1 <- like("cp",
            formula = form1,
            data = datviv,
            samplers = poly,
            domain = dom1)
```

```
lik2 <- like("cp",
            formula = form2,
            data = datcoll,
            samplers = lignes,
            domain = dom2)
```

#### #### ESTIMATION DE PARAMETRE : modélisation conjointe des vivants et des carcasses

```
debut=Sys.time()
jfit <- bru(cmp, lik1, lik2, options = list(control.inla = list(int.strategy = "grid", strategy =
"simplified.laplace")))
fin=Sys.time()
fin-debut
save(jfit,mesh,datviv_train,datcoll_train,
file=paste0("output/rapport_24082022_v2_INLA_monospe30_unstationary_mesh",size_mesh,"_",curr
entID_sp,gsub(" ","",currentNAME_sp),"_12h58_simplifiedlapl_grid.rda"))

summary(jfit)
jfit$summary.fixed
```

#### **Collisions Opportunistes et Collisions Protocolées liées aux Infrastructures de Transport Terrestres**

Déterminants des collisions routières de la faune sauvage et modélisation spatiale intégrée de données protocolées et opportunistes rapport d'étude



**RÉPUBLIQUE  
FRANÇAISE**

*Liberté  
Égalité  
Fraternité*



**Cerema**

CLIMAT & TERRITOIRES DE DEMAIN

# 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

---

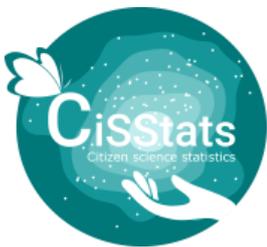
Centre d'Études et d'expertise sur les Risques, l'Environnement, la Mobilité et l'Aménagement

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



UMR 5175  
CENTRE D'ÉCOLOGIE  
FONCTIONNELLE  
& ÉVOLUTIVE



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.

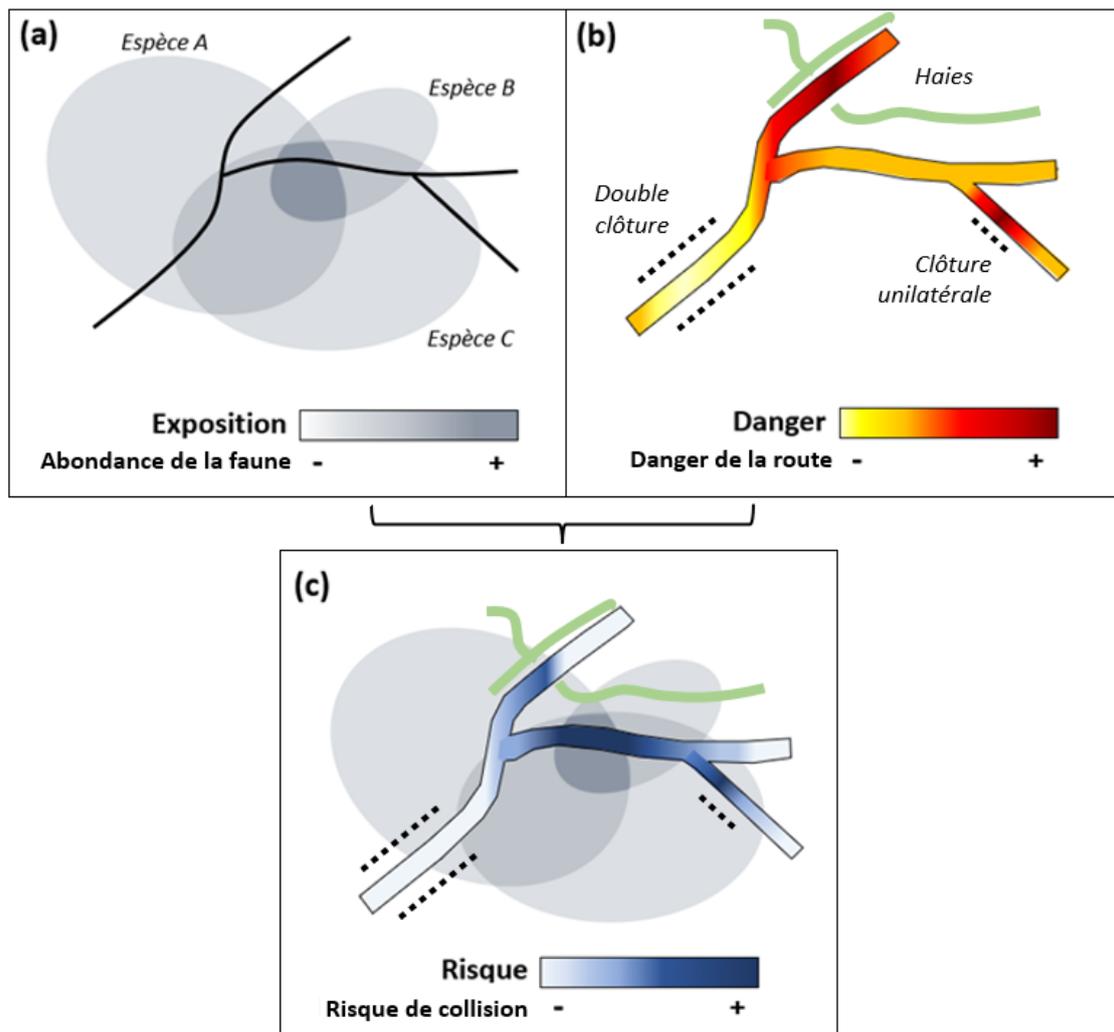


Figure 1 : Schéma de principe d'une analyse de risque couplant l'exposition (a), placement d'un animal de manière qui les expose aux effets de certains équipement, exprimée comme le recouplement de la distribution d'une espèce avec un réseau routier, et le danger (b), propriété intrinsèque d'un équipement 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

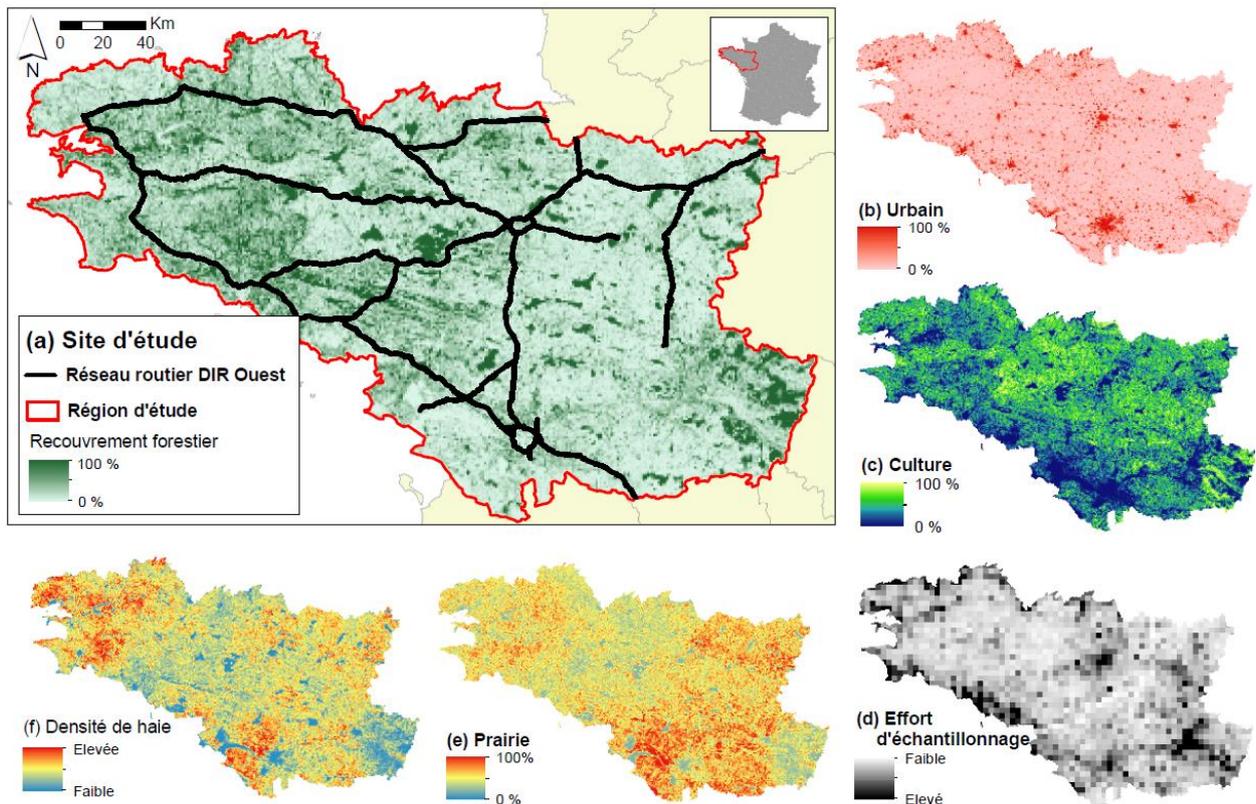


Figure 2 : Site d'étude et son paysage : (a) Région d'acquisition des données, réseau routier de la DIR Ouest et cartographie du recouvrement forestier par  $\text{km}^2$  (b) Recouvrement des zones urbaines par  $\text{km}^2$  (c) Recouvrement des cultures par  $\text{km}^2$  (d) Cartographie de l'effort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

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](http://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 de 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  $\lambda$  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  $\alpha$  et  $\beta$  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  $\alpha$  et  $\beta$  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  $\alpha$  et  $\beta$  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 25km<sup>2</sup>. 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  $\mu$  est formée de composantes latentes gaussiennes. Cette condition est vérifiée si :

- la moyenne  $\mu$  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  $\lambda$  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  $\mu$  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 km<sup>2</sup> : multiplication de l'intensité du centroïde du pixel par l'aire d'un pixel – Intensité de cadavre dans un pixel de 24km<sup>2</sup> : 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 km<sup>2</sup> 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

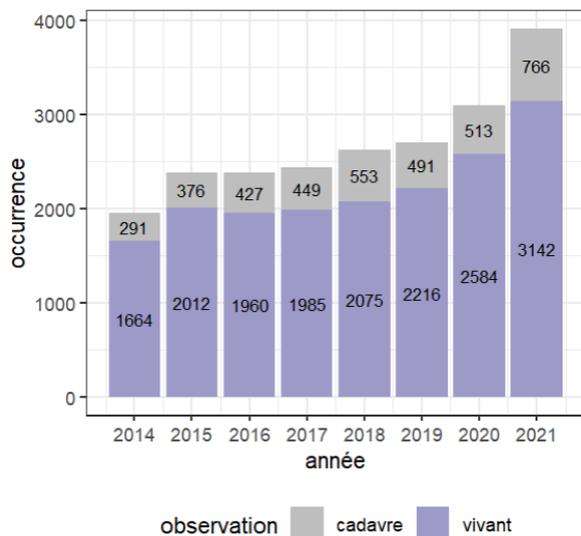


Figure 3: Variation annuelle de l'occurrence des chevreuils vivants sur la région d'étude et de cadavre sur le réseau de la DIR Ouest

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 24km<sup>2</sup> 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 24km<sup>2</sup>. Les zones les plus mortifères se situent au Nord de Rennes, sur les routes partant de Lorient (Fig. 4.b).

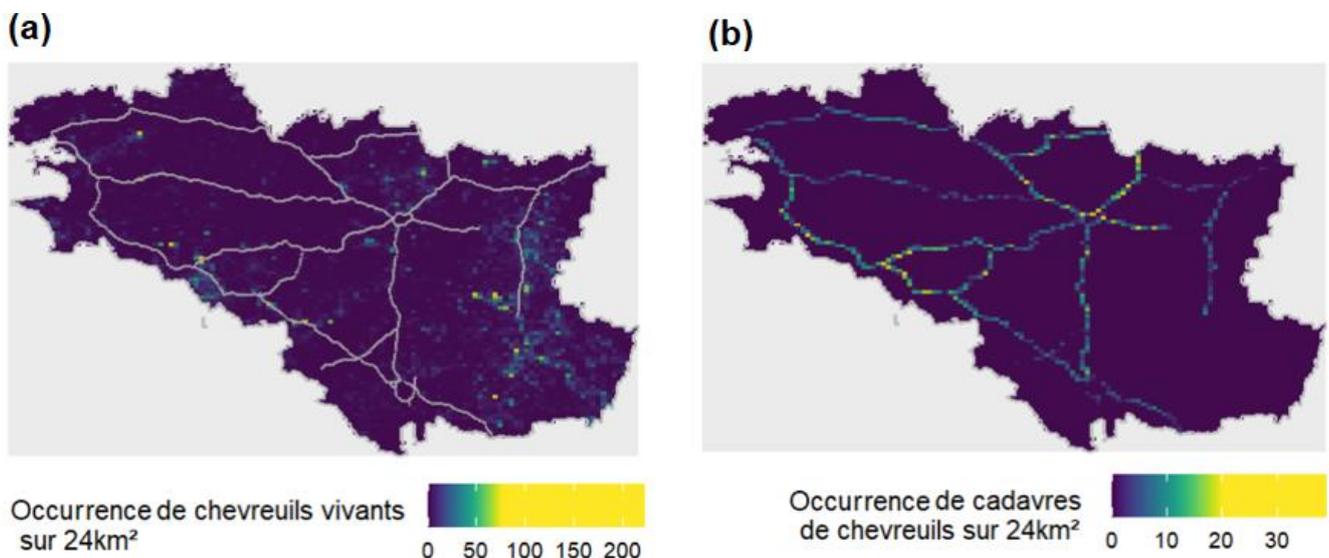
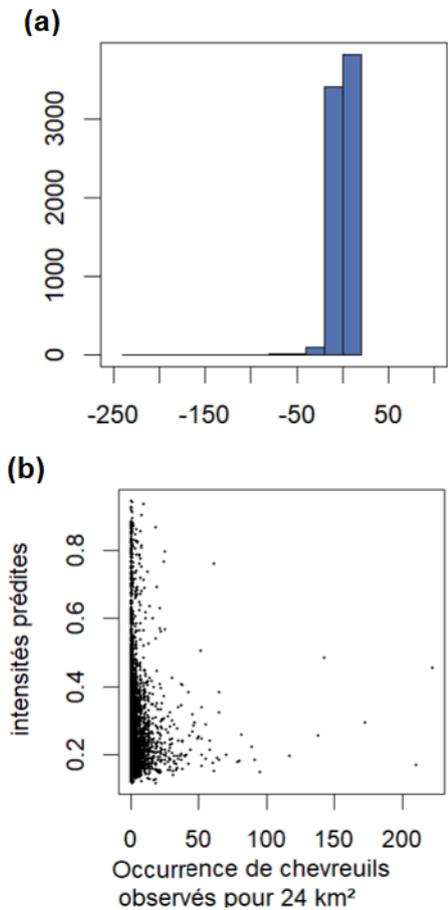


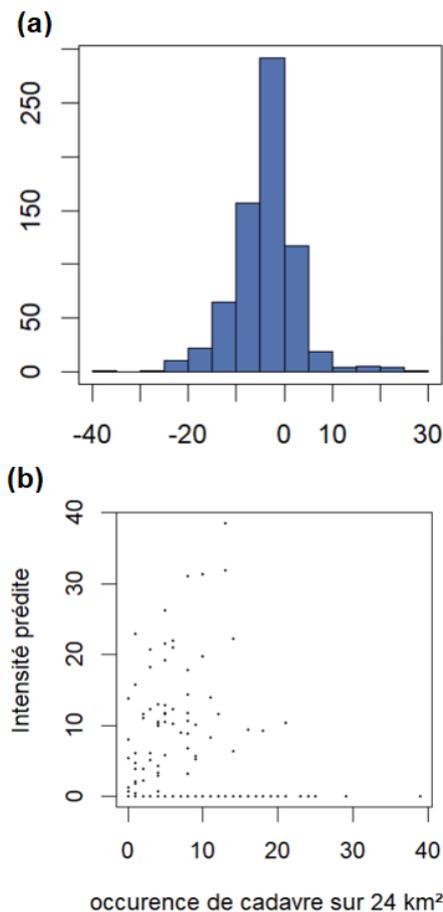
Figure 4 : Observations des chevreuils sur la période de 2014 à 2021, (a) vivants par par le programme Vlsionature, (b) à l'état de cadavre sur le réseau de la DIR Ouest par les patrouilleurs

### 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 24km<sup>2</sup> 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 km<sup>2</sup>) (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).

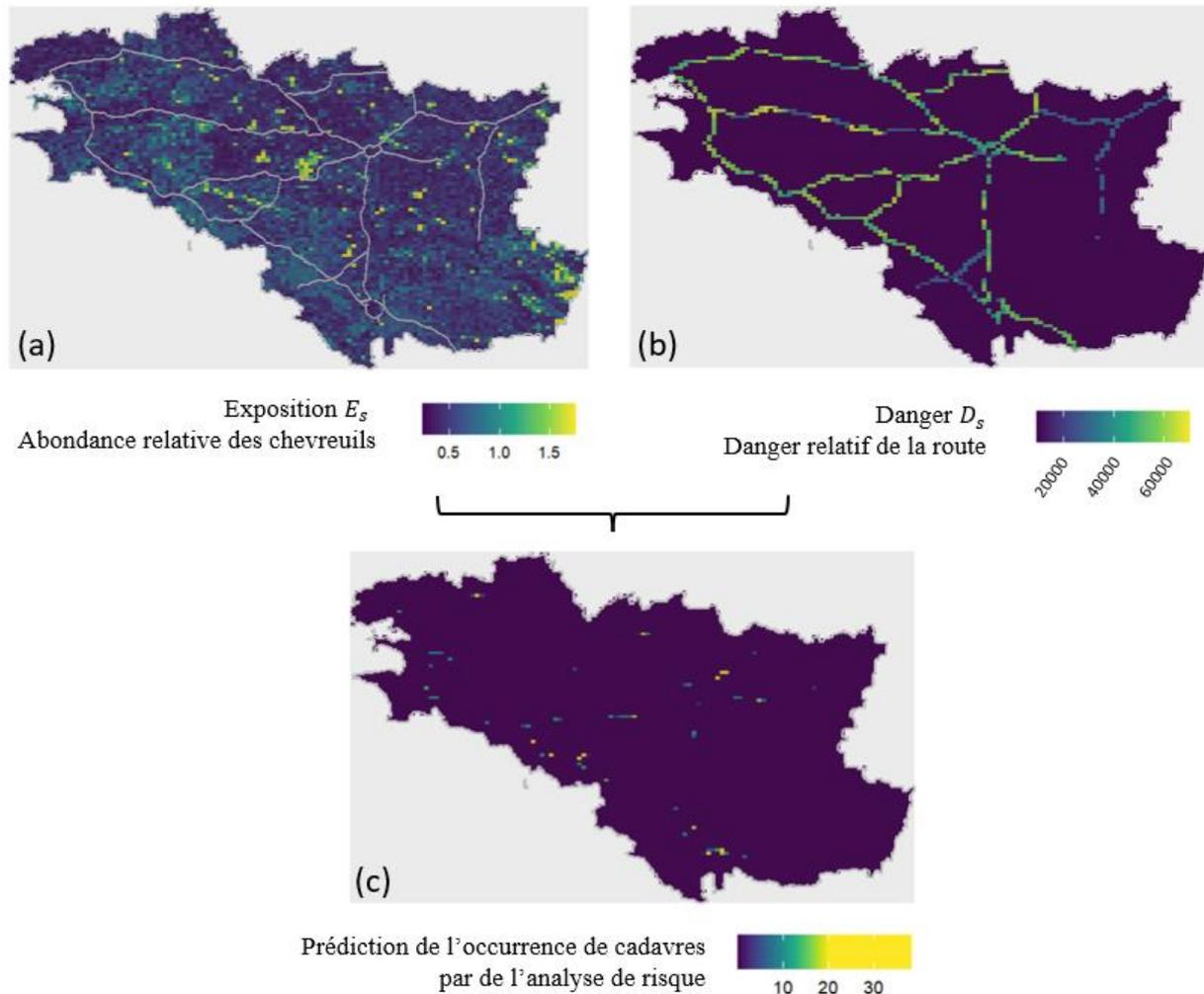


Figure 7 : Prédiction des collisions chevreuil-véhicule par l'analyse de risque entre 2014 et 2021 à résolution de  $24\text{km}^2$ , (a) Prédiction de l'exposition  $E_s$  soit l'abondance relative de chevreuil déterminée par le modèle de niche, (b) Prédiction du danger  $D_s$  intrinsèque à la route, (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  $90\text{km/h}$ . 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).

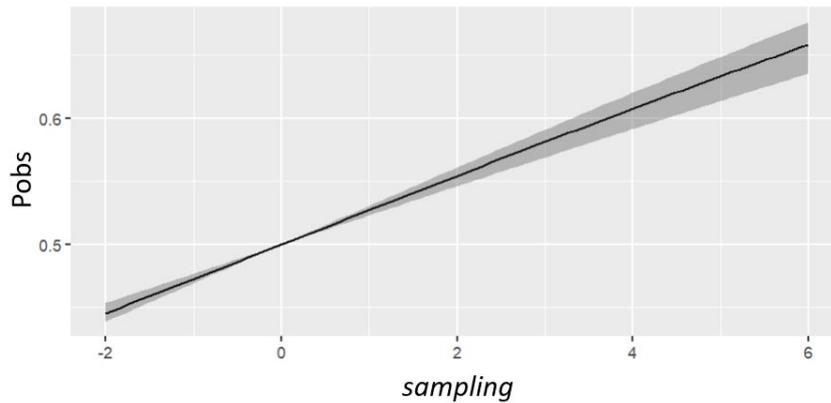


Figure 8 : Relation entre la probabilité d'observer un chevreuil vivant et la covariable d'échantillonnage *sampling*

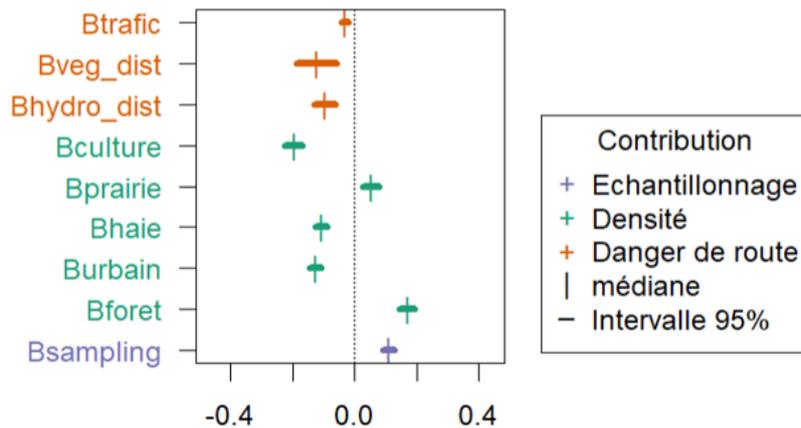


Figure 9 : Estimation des pentes des effet linéaires des covariables sur le danger de la route, la densité relative des chevreuils (exposition) et l'échantillonnage (les B correspondent à  $\beta$ )

- Pour les paramètres d'exposition, l'intercept  $\alpha$  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

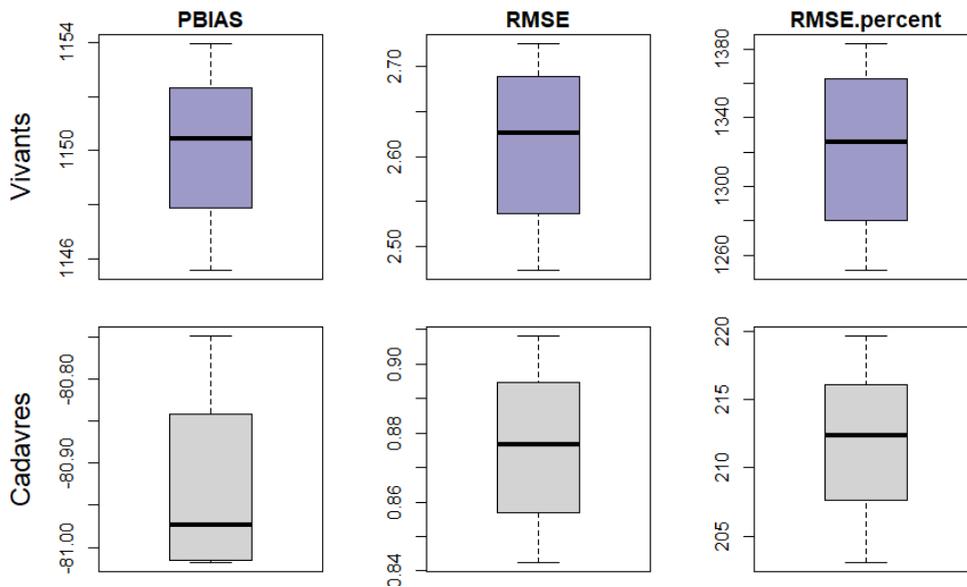


Figure 10 : Indices de la capacité prédictive du modèle dans le temps

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 km<sup>2</sup> représente 211.84% de la variabilité (Fig. 10).

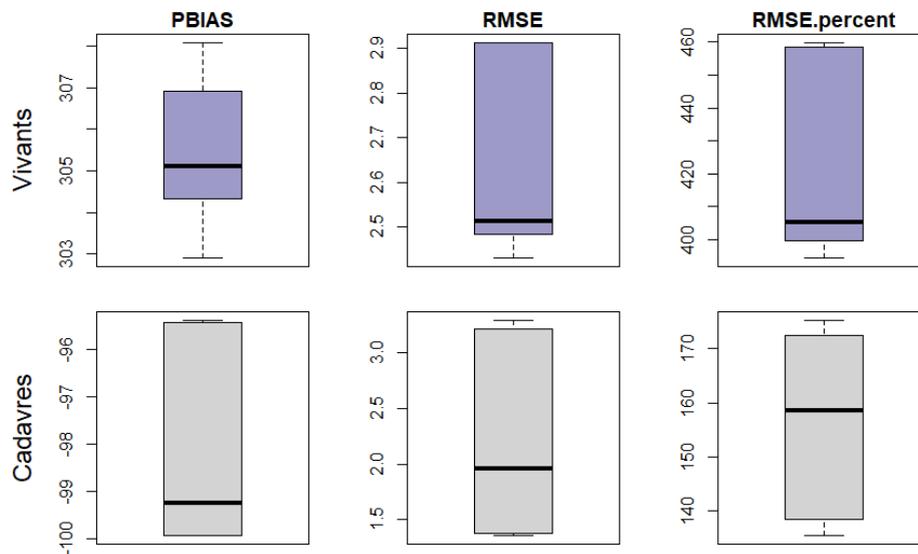


Figure 11 : Indices de la capacité prédictive spatiale du modèle

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 km<sup>2</sup> 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**

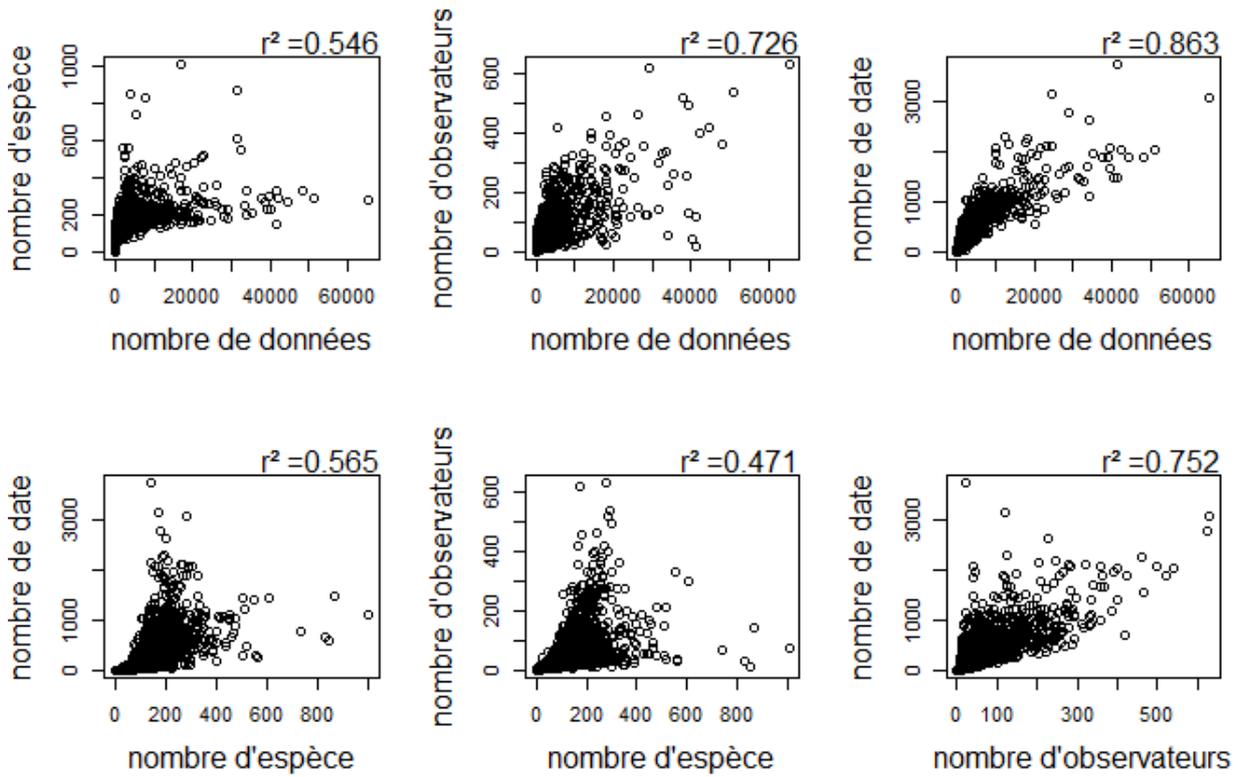
Je tiens à remercier le Cerema qui a financé ce projet ainsi que mes encadrants de stage, Jean-Yves Barnagaud, Jean-François Bretaud et Alain Morand pour leur accueil, leurs conseils, et les nombreux échanges qui ont ponctué ce stage.

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.

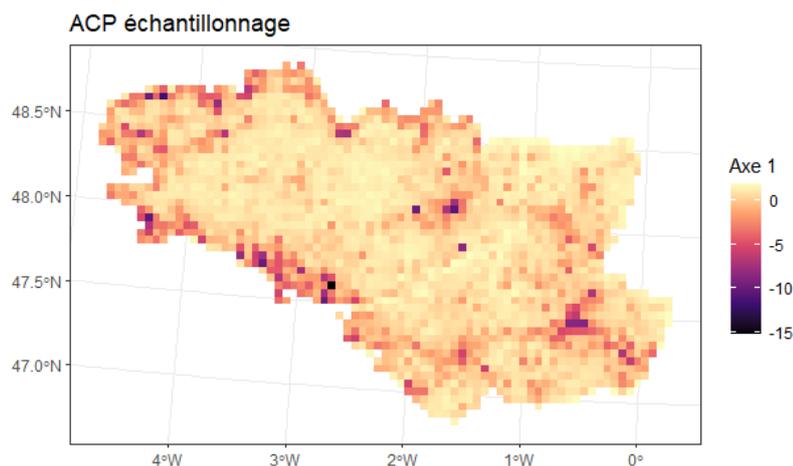
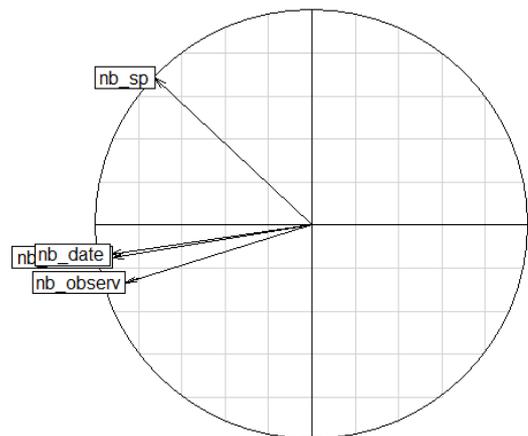
Je remercie chaleureusement l'équipe de l'INRAE BioSP à Avignon qui m'a accueillie l'espace de quelques jours et notamment Julien Papaix qui m'a énormément aidée à la mise en forme ainsi qu'à la compréhension du modèle ainsi que ses explications sur les statistiques spatiales.

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





# COCPITT : Collisions opportunistes et collisions protocolées liées aux Infrastructures de transport terrestre :

## Modélisation spatiale des collisions routières de la faune sauvage en Bretagne

Maëliiss Vandroux – maelissvandroux@gmail.com

J.F. Bretaud, A. Morand, J. Papaix, F. Simonnet, B. Marchadour, E. Guinard, O. Pichard, J.Y. Barnagaud



AGIR pour la  
BIODIVERSITÉ  
PAYS DE LA LOIRE



École Pratique  
des Hautes Études



AGIR pour la  
BIODIVERSITÉ  
BRETAGNE



1

## Sommaire

- Introduction
- Réflexion sur le modèle
- Modèle préliminaire
- Modéliser l'exposition
- Modéliser le danger routier
- Modéliser le risque de collision
- Conclusion
- Perspectives

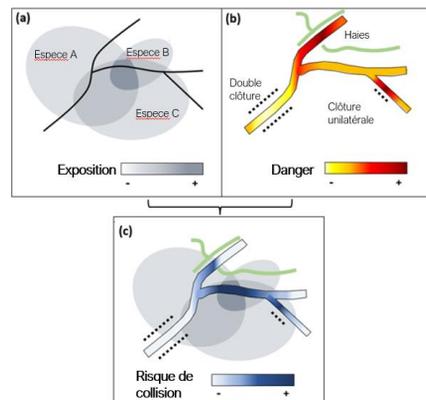


Schéma du modèle préliminaire

2

2

## Un enjeu de conservation et de sécurité publique

### Collisions



Le long des routes nationales de Bretagne :

> **5000** collisions / an pour 1500 km

Ce nombre est sous-estimé à minima par un facteur 5 (projet exploratoire COMERCAR ITTECOP)

3

3

## Un enjeu de conservation et de sécurité publique

### Collisions



Le long des routes nationales de Bretagne :

> **5000** collisions / an pour 1500 km

Ce nombre est sous-estimé à minima par un facteur 5 (projet exploratoire COMERCAR ITTECOP)

### Atténuation



Besoin d'une **prediction spatiale** du risque de mortalité par collision à échelle régionale.



4

4

## Comment modéliser la mortalité par collision ?

Par quel processus ?

Particularités de détection ?

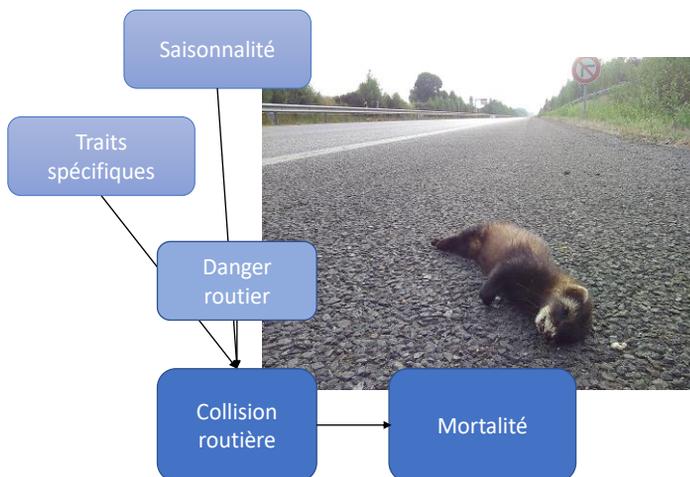


5

## Comment modéliser la mortalité par collision ?

Par quel processus ?

Particularités de détection ?



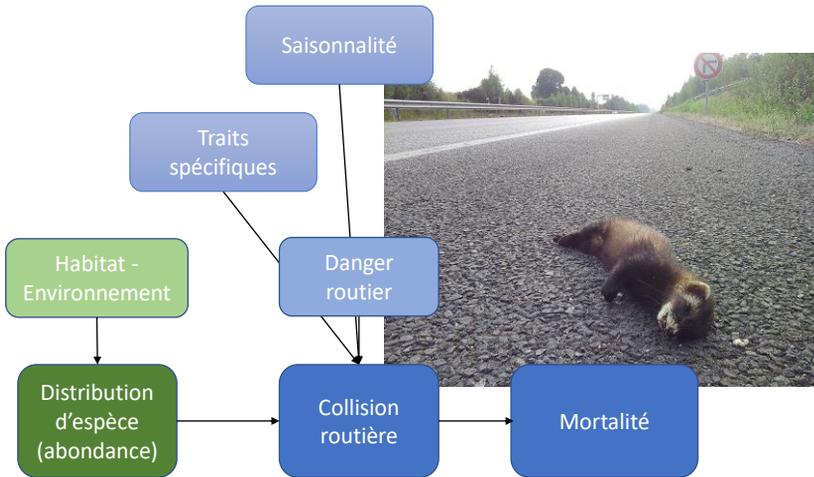
6

Réflexion sur le modèle

## Comment modéliser la mortalité par collision ?

Par quel processus ?

Particularités de détection ?



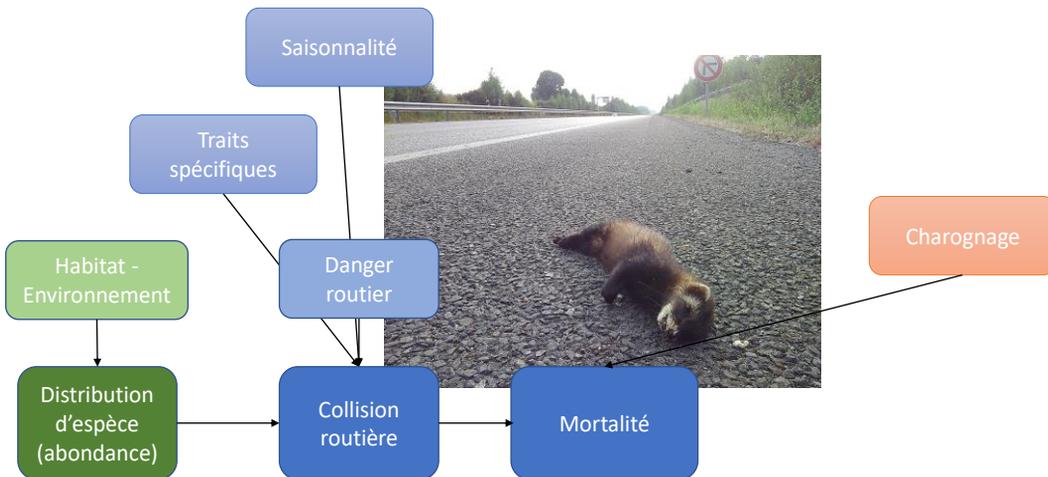
7

Réflexion sur le modèle

## Comment modéliser la mortalité par collision ?

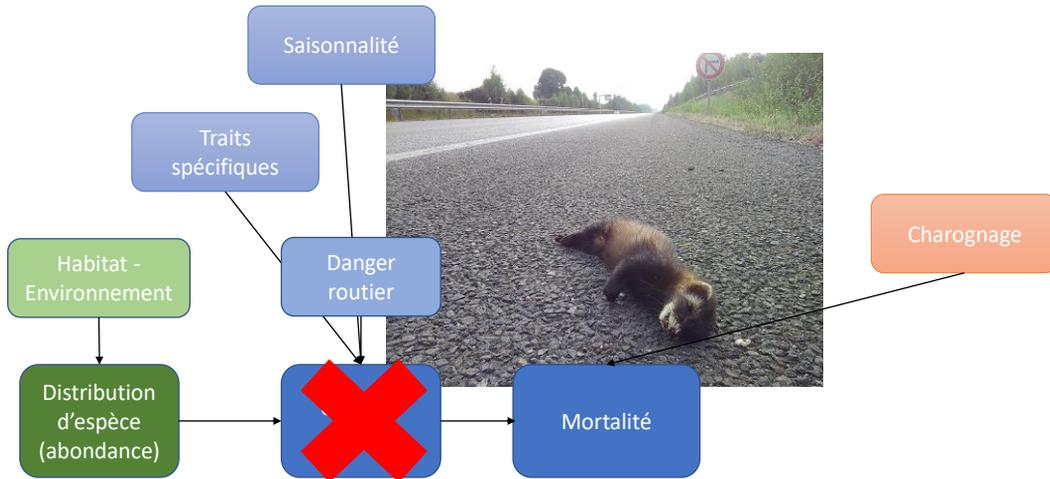
Par quel processus ?

Particularités de détection ?



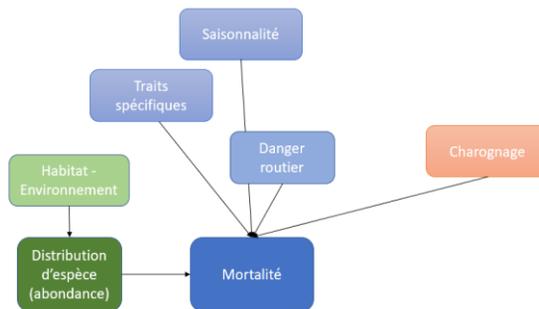
8

## Comment modéliser la mortalité par collision ?



9

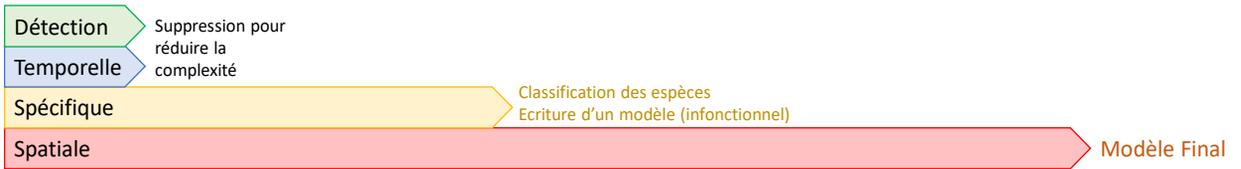
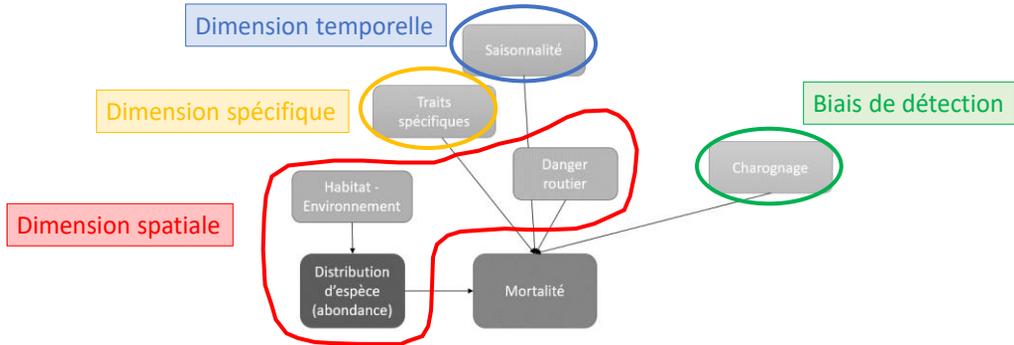
## Comment modéliser la mortalité par collision ?



10

Réflexion sur le modèle

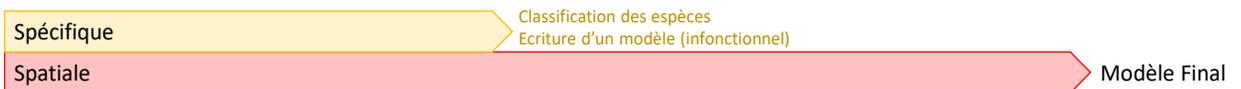
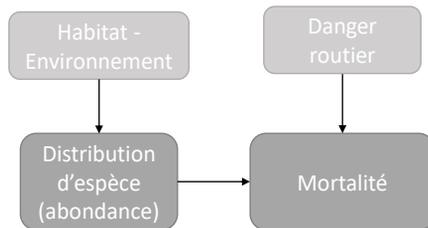
Quelle dimension ?



11

Réflexion sur le modèle

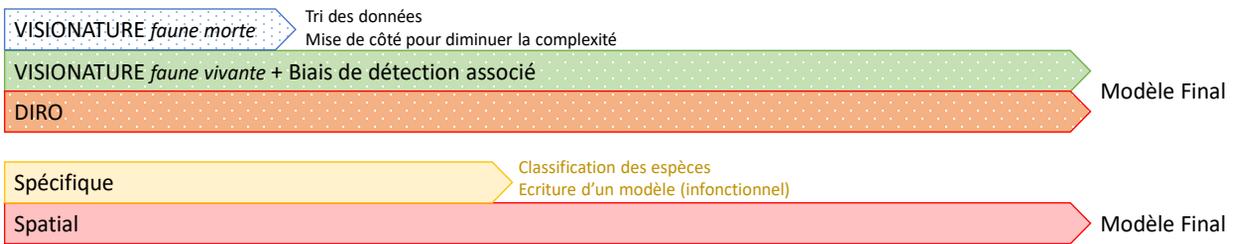
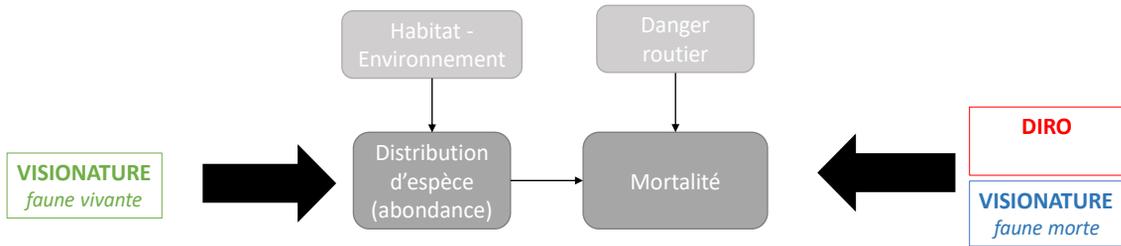
Quelle dimension ?



12

Réflexion sur le modèle

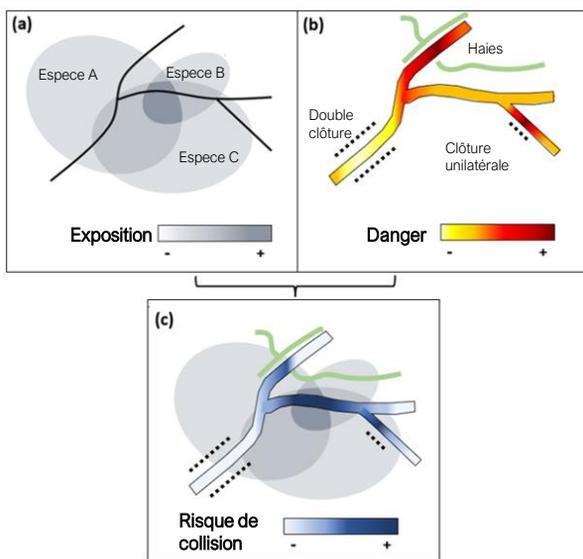
Quelles données ?



13

Modèle préliminaire

Un analyse de risque



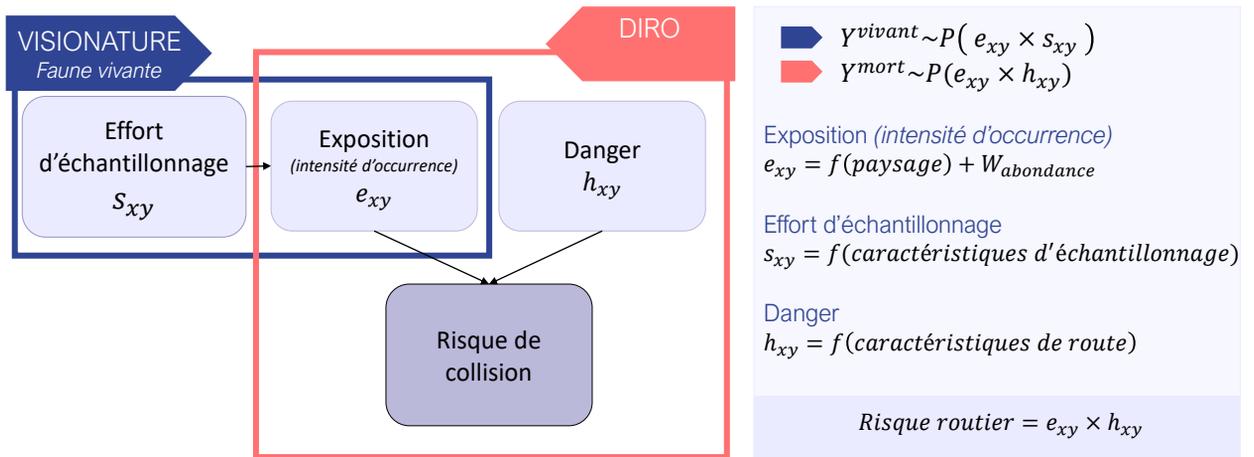
Définitions

- (a) **Exposition:** Placement d'un animal de manière qui l'expose aux routes = où l'animal est positionné par rapport aux routes
- (b) **Danger:** Propriété intrinsèque des routes d'avoir un effet dangereux = où les collisions arrivent le long des routes

14

14

## Un modèle joint



15

15

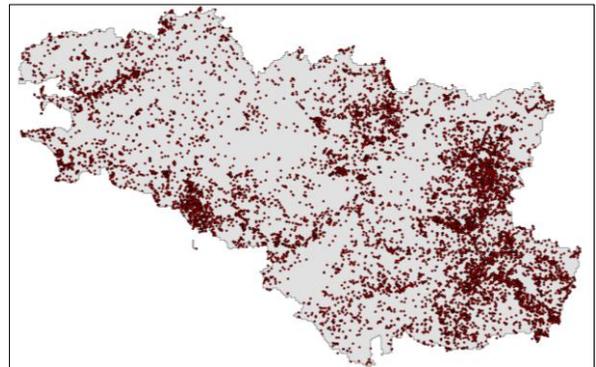
## Données de chevreuil vivants: Des données opportunistes

## Avantages

- Données disponibles pour la région
- Pour une grande période (8 ans)
- Beaucoup de données (17 208 points)

## Limites

- Effort d'échantillonnage hétérogène dans l'espace et dans le temps



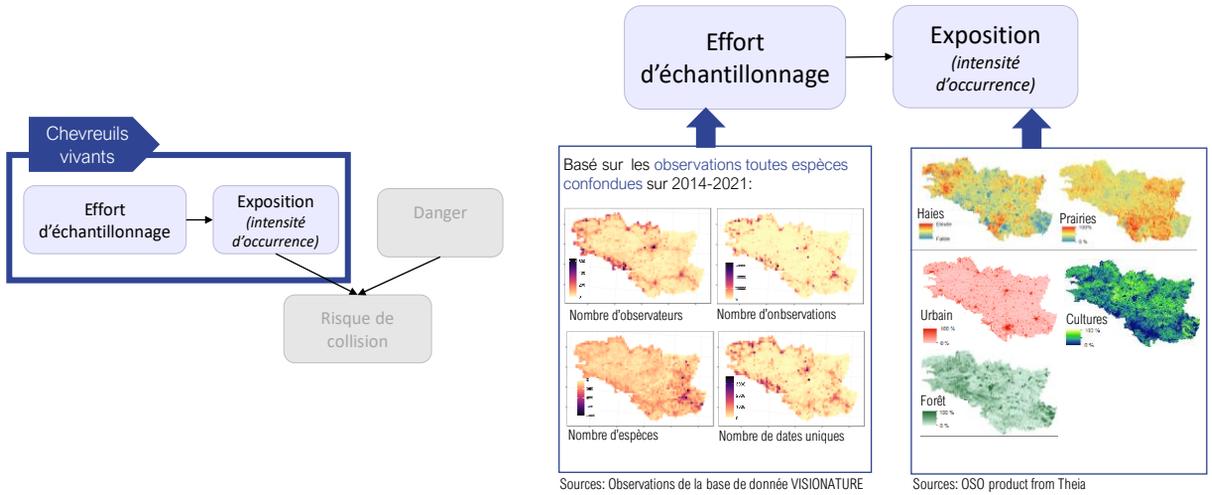
Observations de chevreuils vivants de la base de données VISIONATURE entre 2014 et 2021

16

16

Modéliser l'exposition

# Prédicteurs



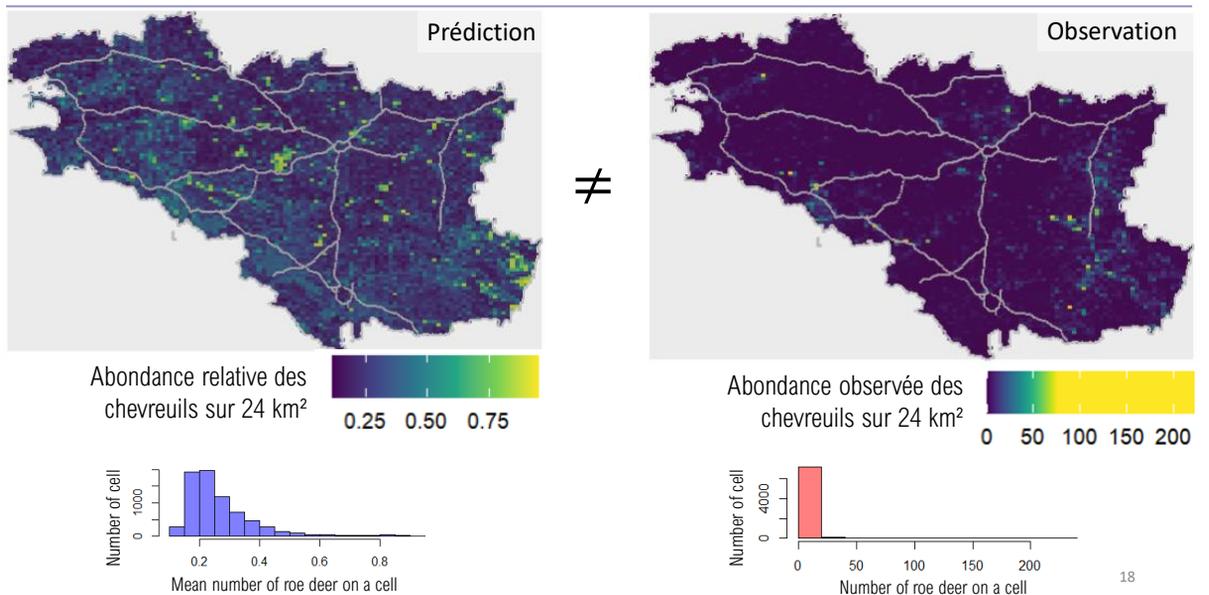
De Solan, T., Renner, I., Cheylan, M., Geniez, P. & Barnagaud, J.-Y. (2019) Opportunistic records reveal Mediterranean reptiles' scale-dependent responses to anthropogenic land use. *Ecography*

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*

17

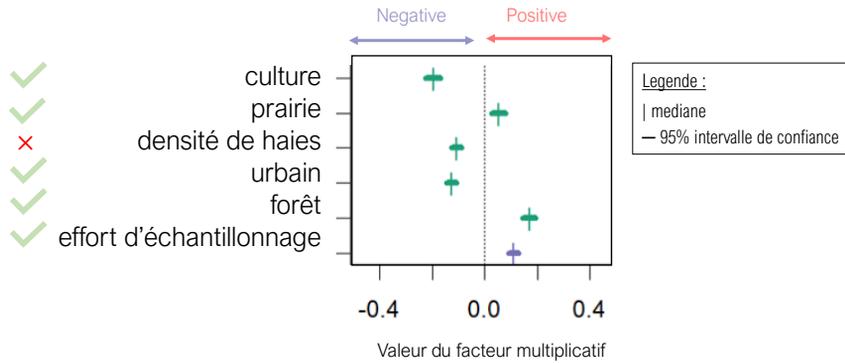
Modéliser l'exposition

# Premiers résultats : cartographie



18

## Premiers résultats : effet des prédicteurs



19

19

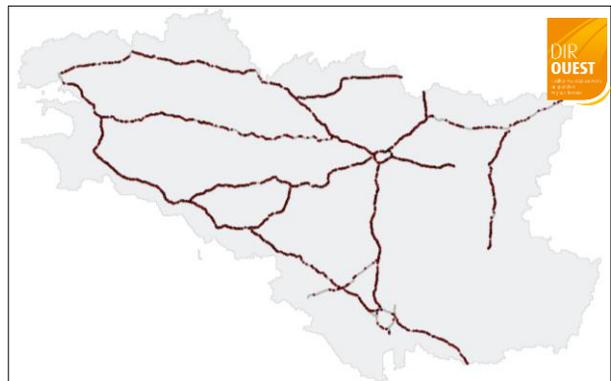
## Données de collisions semi-protocoles

## Avantages

- Données disponibles pour la région
- Pour une grande période (8 years)
- Beaucoup de données (3859 points)
- Prospection quotidienne

## Limites

- Prospection à haute vitesse (~ 80 km/h)
- Prospection par des non-naturalistes
- Tâche subsidiaire



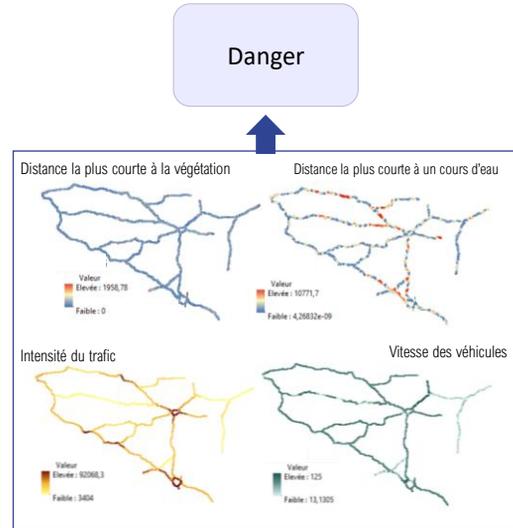
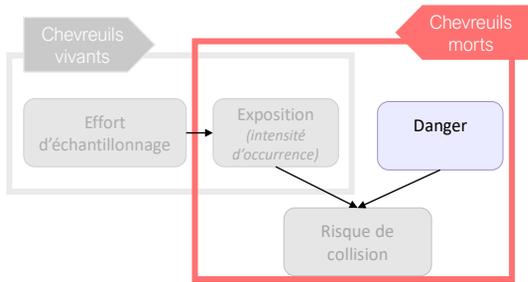
Carcasses de chevreuil détectées par les patrouilleurs de la DIR Ouest entre 2014 et 2021

20

20

Modéliser le danger

Prédicteurs



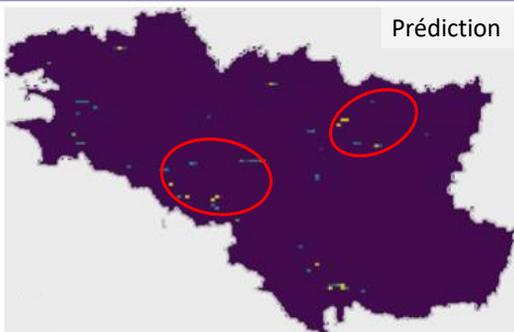
Sources: calculations from IGN data of the BDTOPO

21

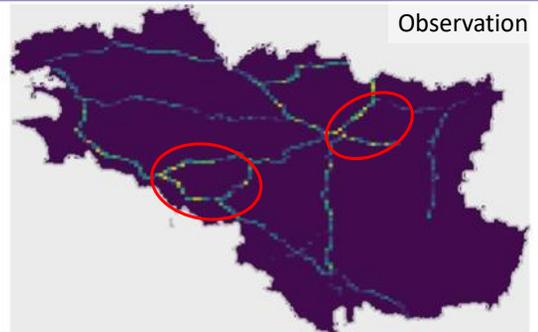
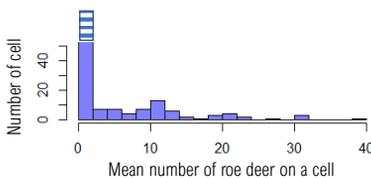
21

Modéliser le risque

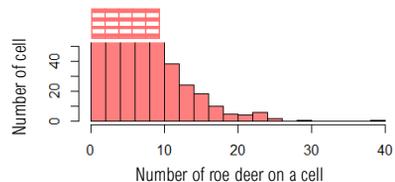
Premiers résultats : cartographie



Nombre simulé de chevreuils morts par collision pour 24km<sup>2</sup>



Nombre observé de chevreuils morts par collision pour 24 km<sup>2</sup>



22

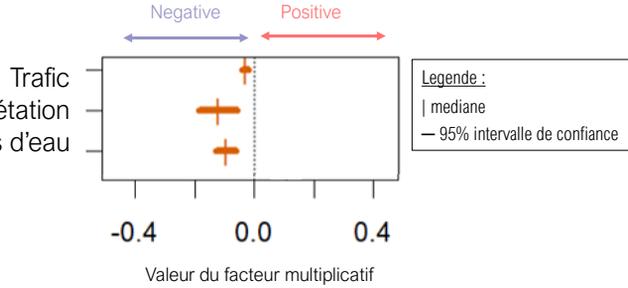
22

Modéliser le risque

Premiers résultats : effet des prédicteurs du danger



- Plus proche distance à la végétation
- Plus proche distance à un cours d'eau



Effet d'une vitesse Moyenne < Effet d'une vitesse haute

23

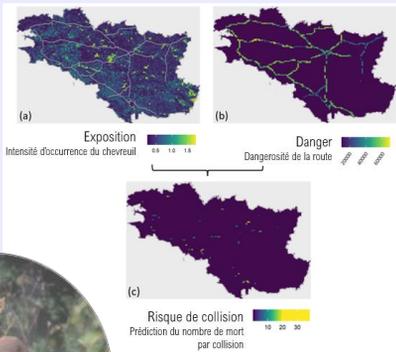
23

Conclusion

Etat actuel du modèle

- OÙ ? ✗
- POURQUOI ? ✓
- PRÉDIRE ? ✗

Insuffisant à des fins d'aménagement du territoire.  
 Résultats encourageants concernant les effets des prédicteurs.



24

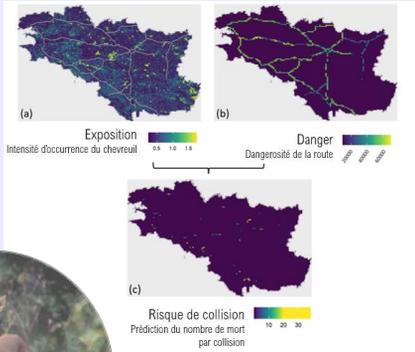
24

Conclusion

Etat actuel du modèle

OÙ ? ✗  
 POURQUOI ? ✓  
 PRÉDIRE ? ✗

Insuffisant à des fins d'aménagement du territoire.  
 Résultats encourageants concernant les effets des prédicteurs.



En réalisation

- Méthode pour établir une carte du risque de collision routière intégrant:
  - la distribution de l'espèce
  - le danger routier
- Méthode pour valoriser des données issues de science participative
- Méthode de regroupement d'espèce en vue d'une analyse multi espèce

Limites

- Modèle mono-espèce
- Modèle sur une espèce commune
- Manque de validation par rapport à des données protocolées

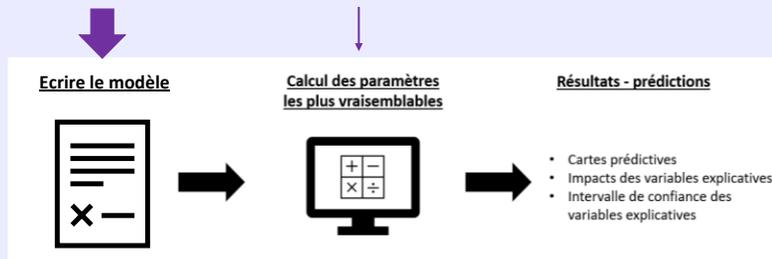
Perspectives

Améliorer l'ajustement du modèle



J. Papaix

- Rajout de champs d'autocorrélation spatiaux
- Rajout de relations non linéaires pour les effets des prédicteurs
- Intégration d'une analyse spécialisée pour les réseaux



A long terme

- Faire un modèle multi-espèce
- Intégrer d'autres sources de données au modèle
- Valider le modèle par rapport à des données protocolées



# Spatial joint modelling of wildlife casualties along major roads of Brittany, France.

Maëliss Vandroux – maelissvandroux@gmail.com

J. Papaix, J.F. Bretaud, A. Morand, F. Lasgorceux, F. Simonnet, B. Marchadour, E. Guinard, O. Pichard, J.Y. Barnagaud



**ITTEGOP**  
Infrastructures  
de transports terrestres  
écosystèmes et paysages



**Cerema**



École Pratique  
des Hautes Études

PSL 



AGIR pour la  
**BIODIVERSITÉ**  
PAYS DE LA LOIRE



Une voix pour la nature



AGIR pour la  
**BIODIVERSITÉ**  
BRETAGNE

# Wildlife collision : an issue of conservation and safety on road

---

Roadkill



Along major roads of Brittany :

> 5000 wildlife casualties / year for 1500 km

These numbers of traffic death are underestimated by at least a 5-fold factor.  
(COMERCAR exploratory project ITTECOP)

Mitigation



# Wildlife collision : an issue of conservation and safety on road

Roadkill



Along major roads of Brittany :

> **5000** wildlife casualties / year for 1500 km

These numbers of traffic death are underestimated by at least a 5-fold factor.  
(COMERCAR exploratory project ITTECOP)

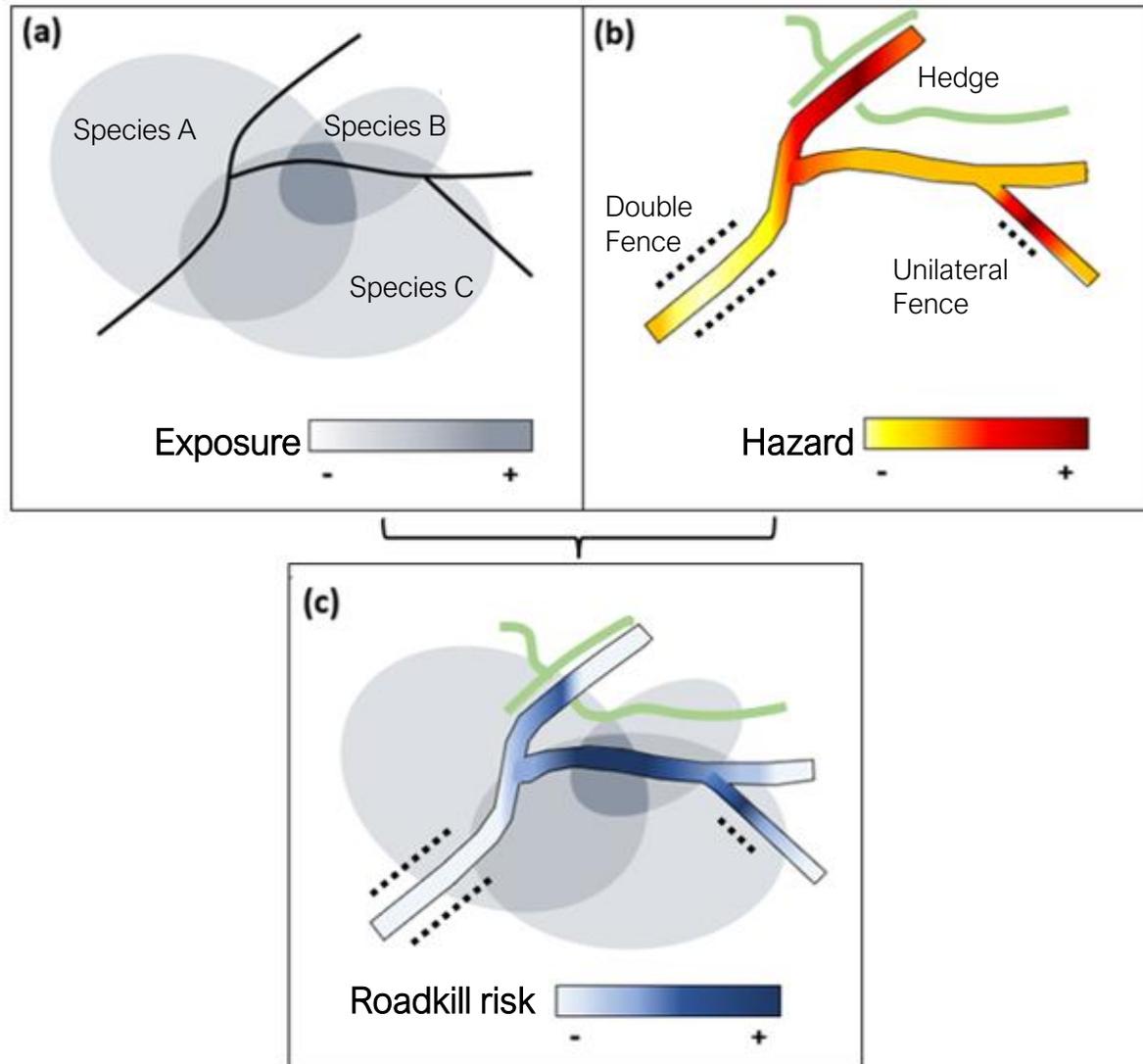
Mitigation



Need for a **spatial prediction** of wildlife roadkill risk at a regional scale.



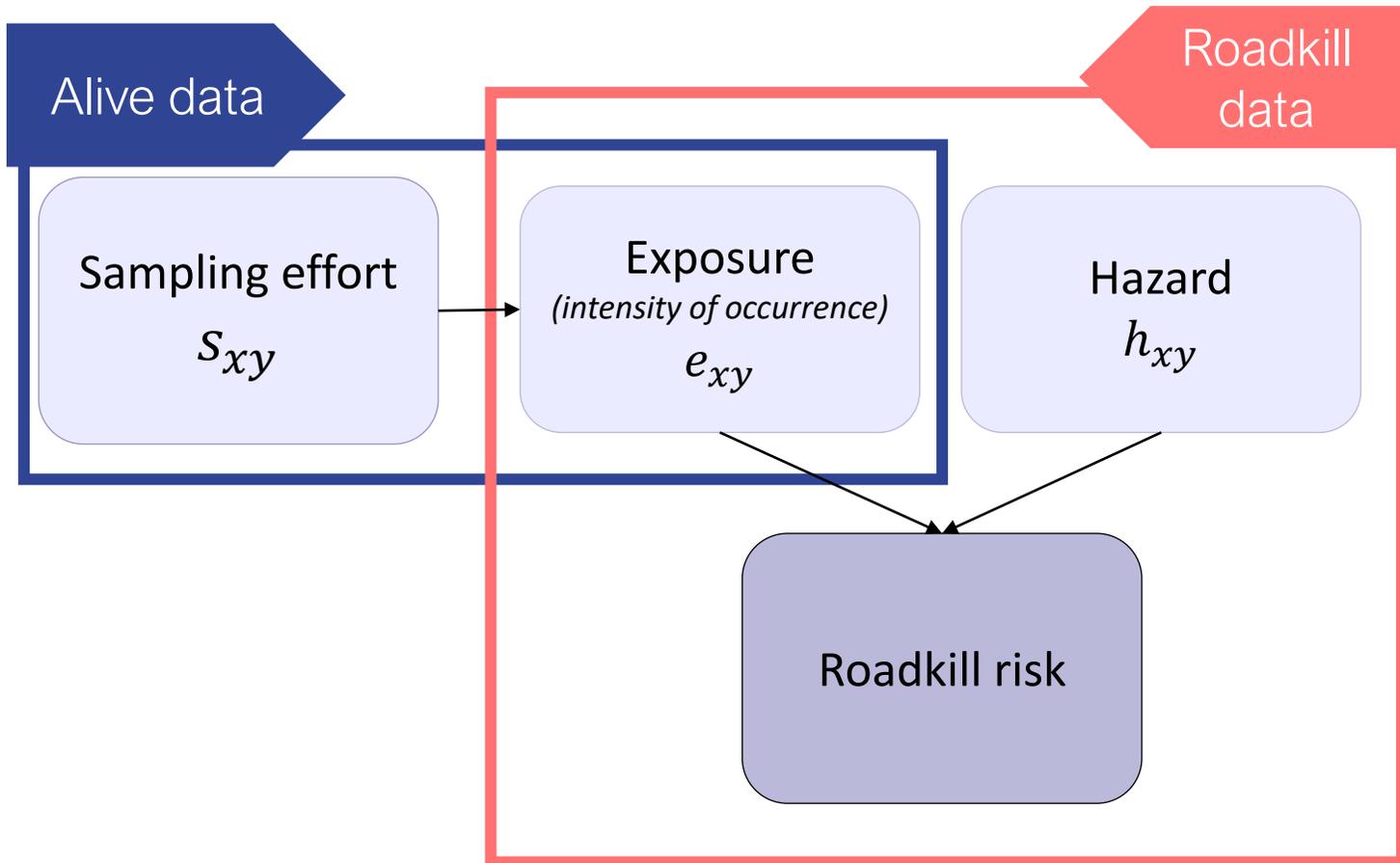
# Modeling wildlife collisions as a risk analysis



## Definitions

- (a) Exposure:** Placing an animal in a manner that exposes it to roads = *where animals are located with respect to roads*
- (b) Hazard:** an intrinsic property of a road that may have a harmful effect = *where collisions occur along roads*

# Modeling Roadkill Risk : a joint model



■  $\gamma^{alive} \sim P(e_{xy} \times s_{xy})$

■  $\gamma^{roadkill} \sim P(e_{xy} \times h_{xy})$

Exposure (intensity of occurrence)  
 $e_{xy} = f(\text{landscape}) + W_{abundance}$

Sampling effort  
 $s_{xy} = f(\text{sampling characteristics})$

Hazard  
 $h_{xy} = f(\text{road characteristics}) + W_{roadkill}$

$$\text{Roadkill risk} = e_{xy} \times h_{xy}$$

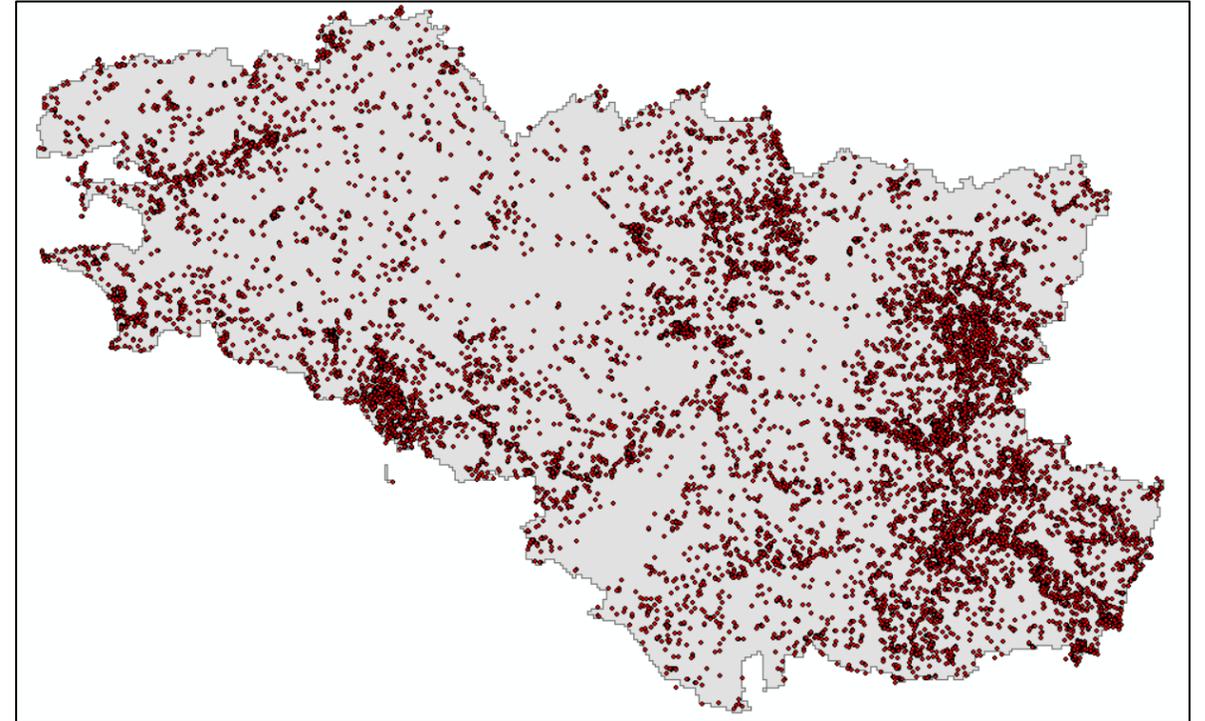
# Alive roe deer data : opportunistic records

## Key virtues

- Data available for a regional extent
- Extended time period (*7 years*)
- Large data set (*17 208 data points*)
- Limited production cost

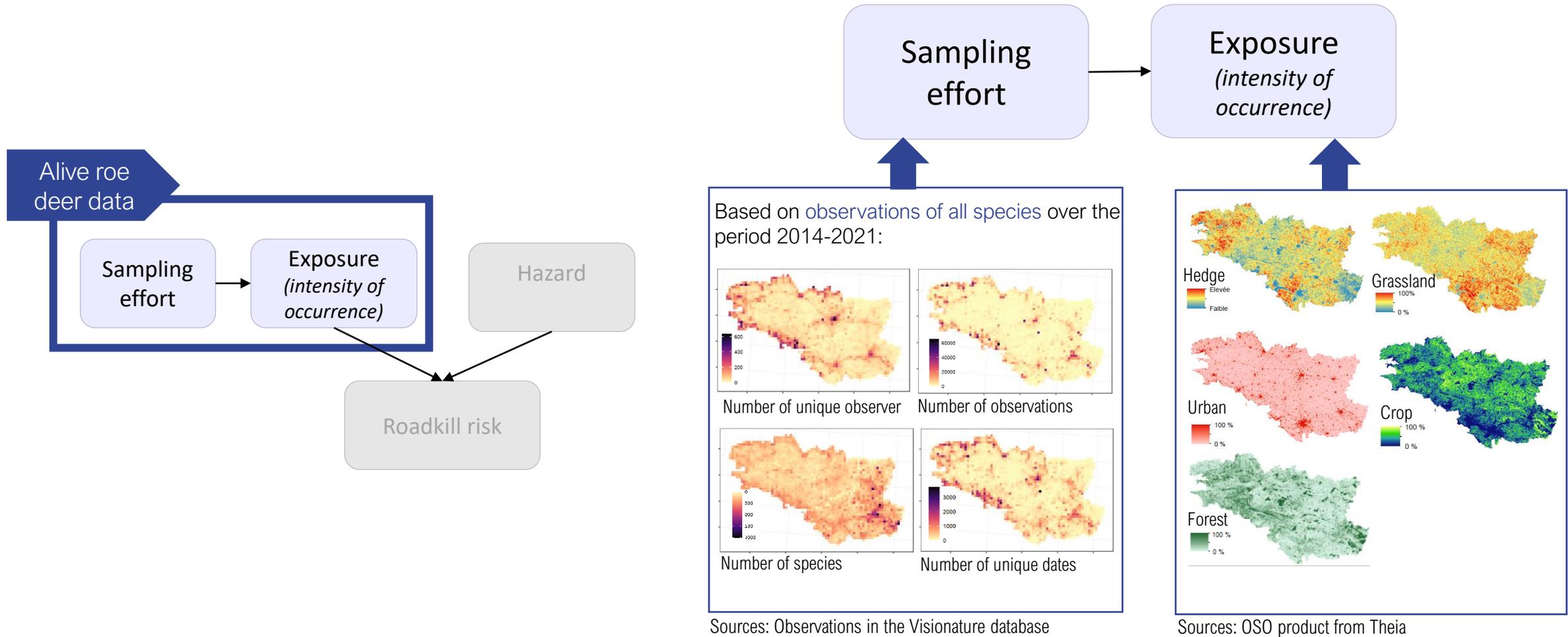
## Key limitations

- Heterogeneous sampling effort in time and space
- Unknown spatial grain

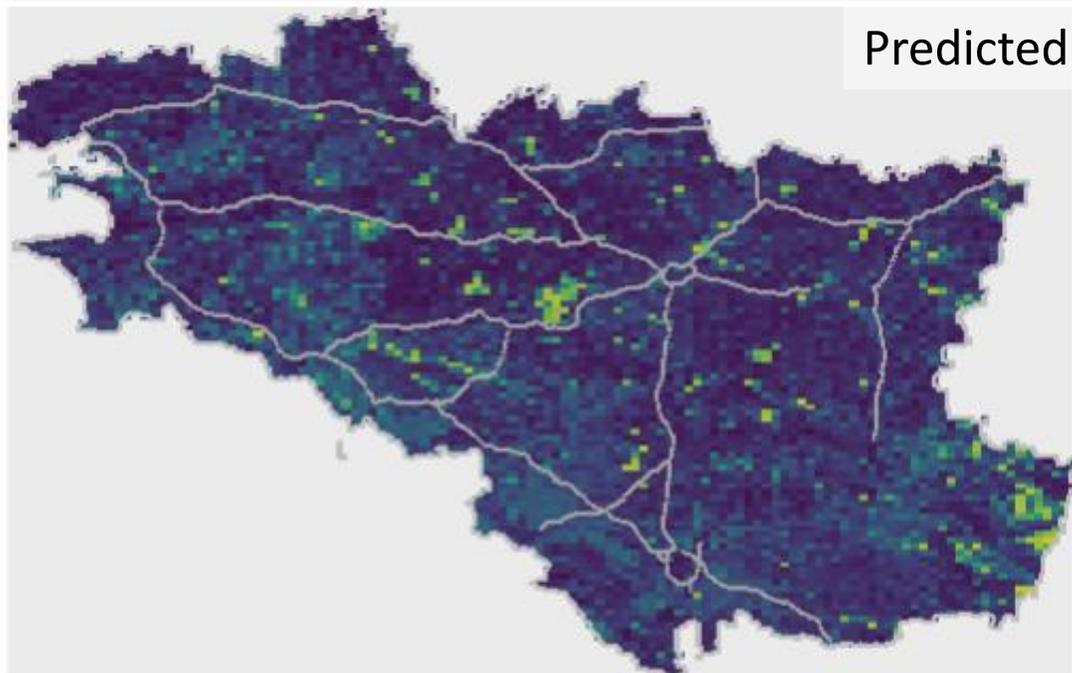


Alive roe deer record points in the Visionature database between 2014 and 2021

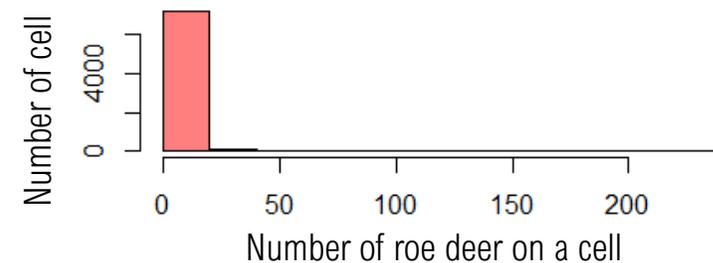
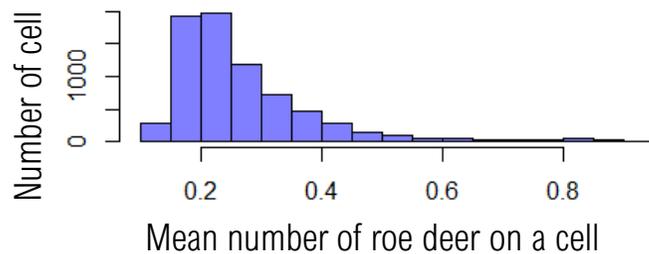
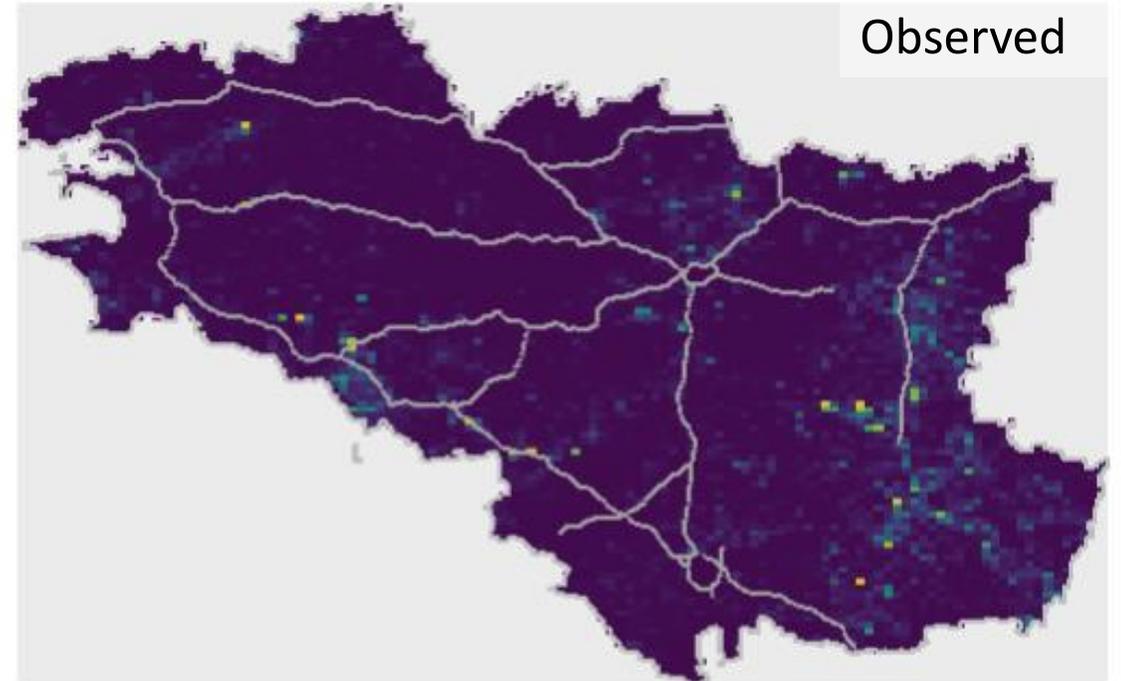
# Modelling exposure



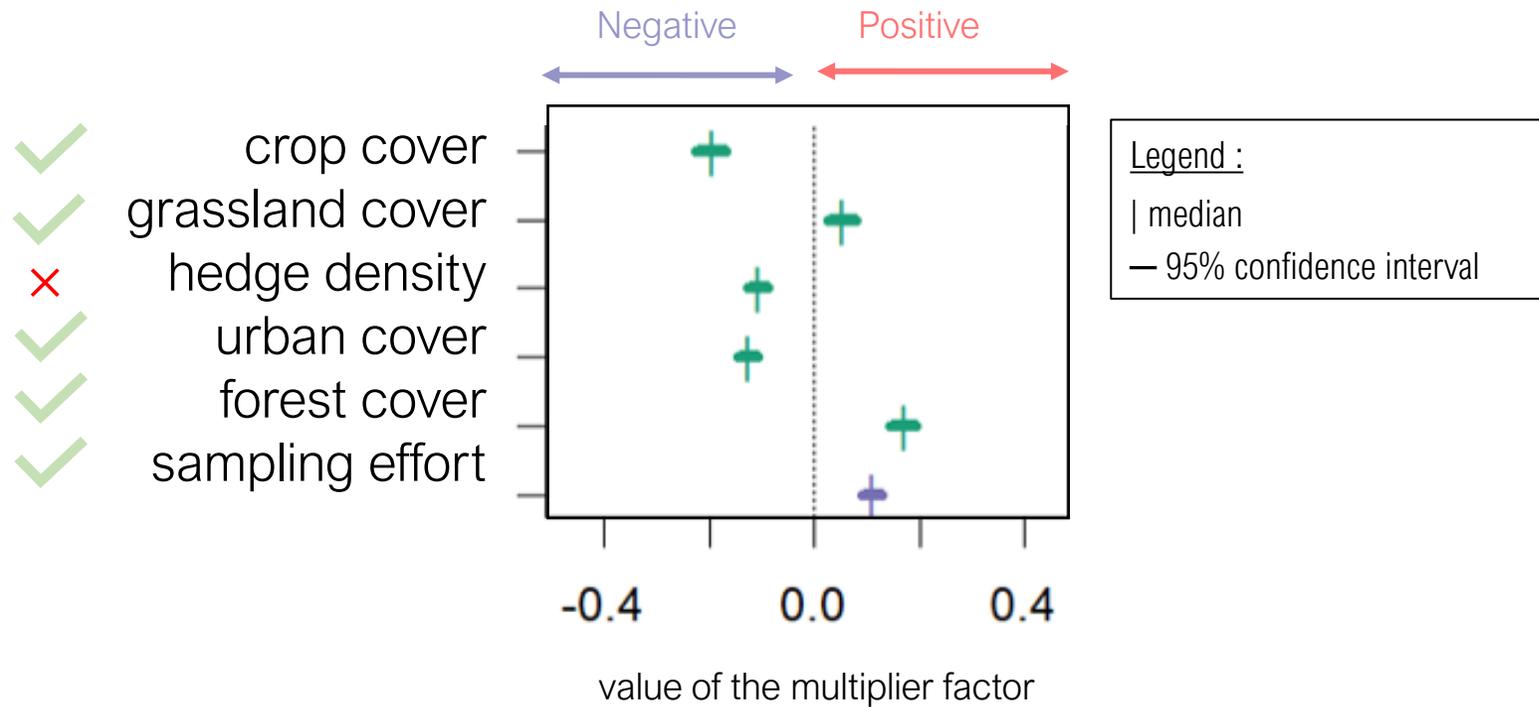
# Modelling exposure : first results



≠



# Modelling exposure : first results



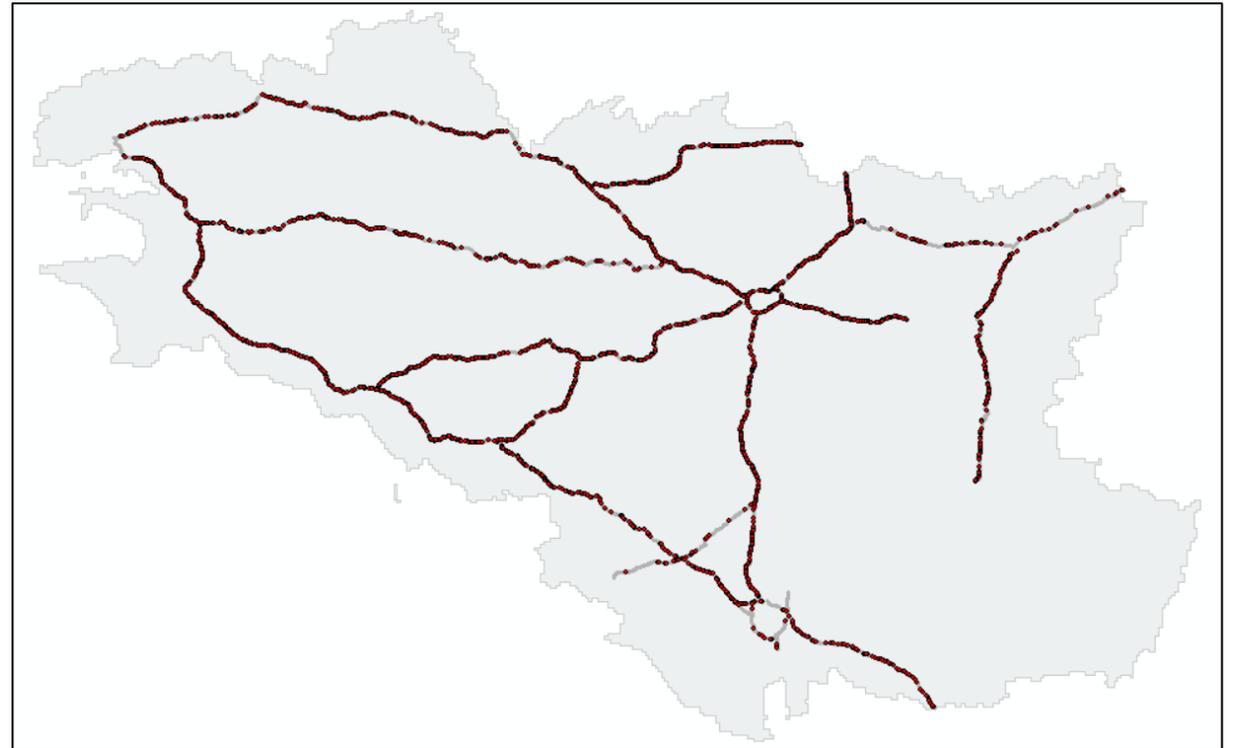
# Roadkill data : semi-protocolled data

## Key virtues

- Data available for a regional extent
- Extended time period (*7 years*)
- Large data set (*3859 data points*)
- Daily survey on the national road network
- Low cost

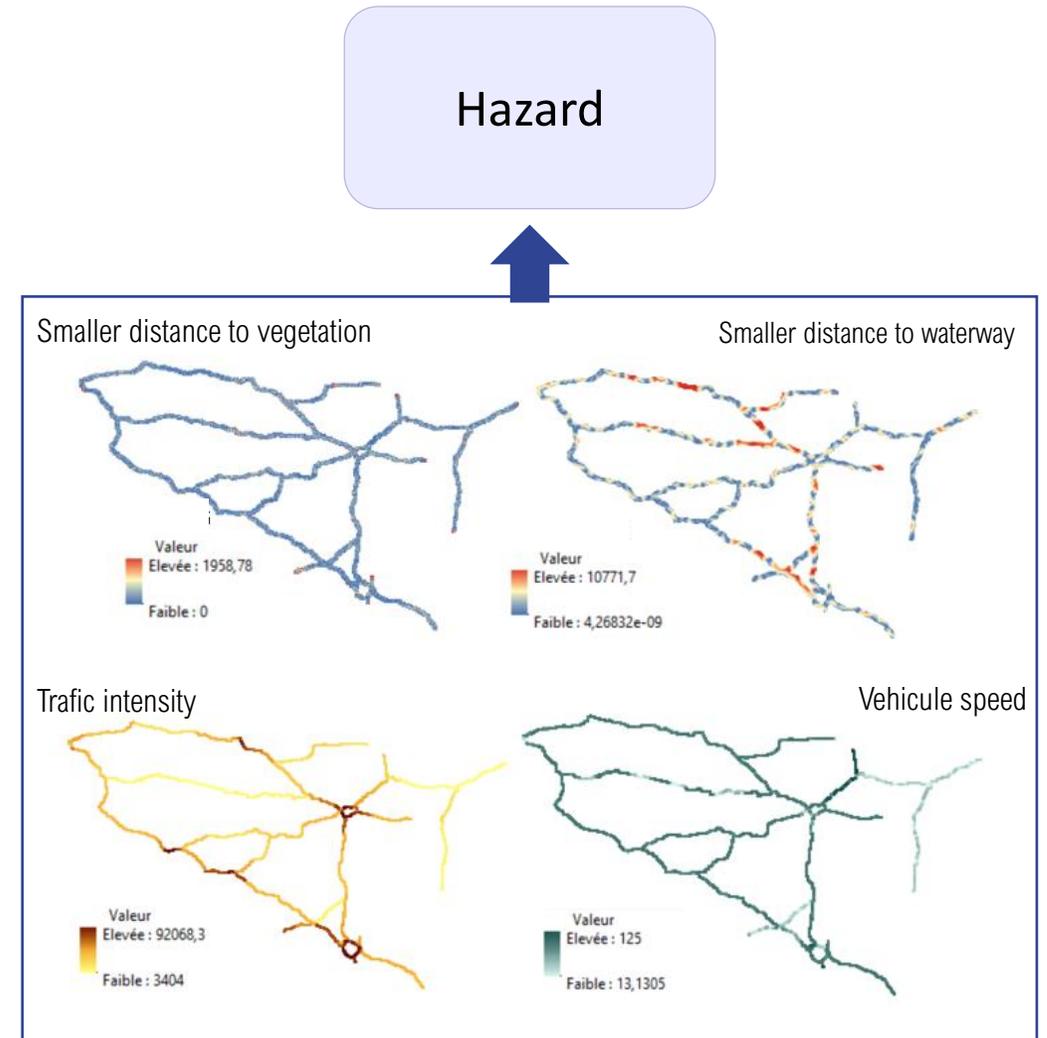
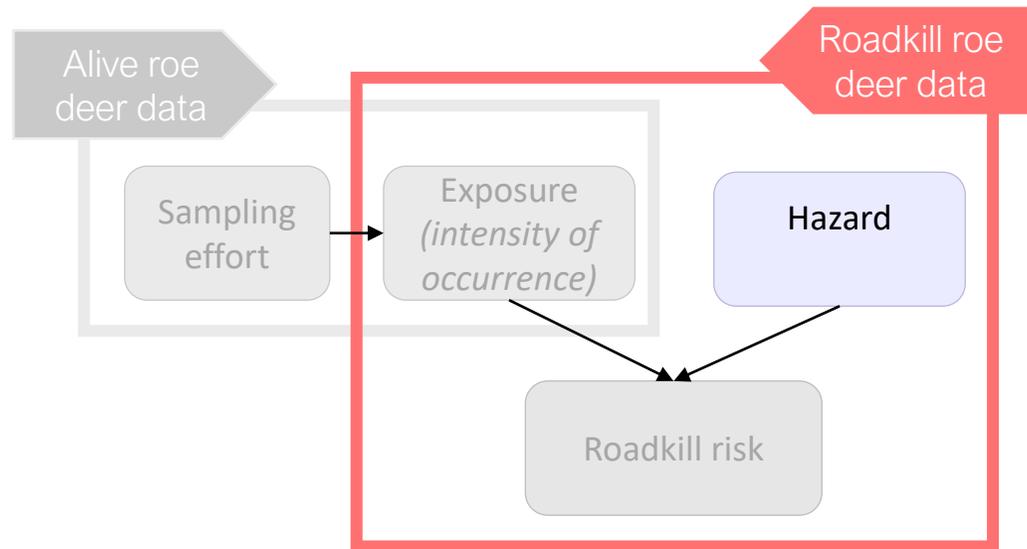
## Key limitations

- Survey at a high speed (60 to 80 km/h)
- Survey by non-naturalists
- Surveying is a subsidiary task



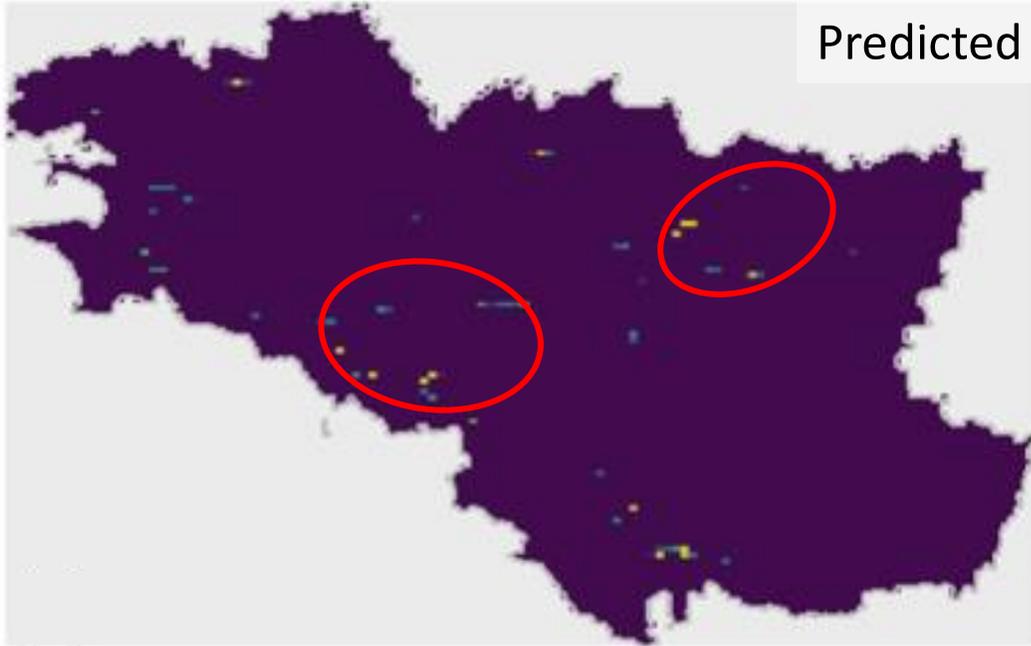
Roe deer carcasses observed by DIR Ovest patrollers during their daily patrols between 2014 and 2021

# Modelling roadkill hazard : method



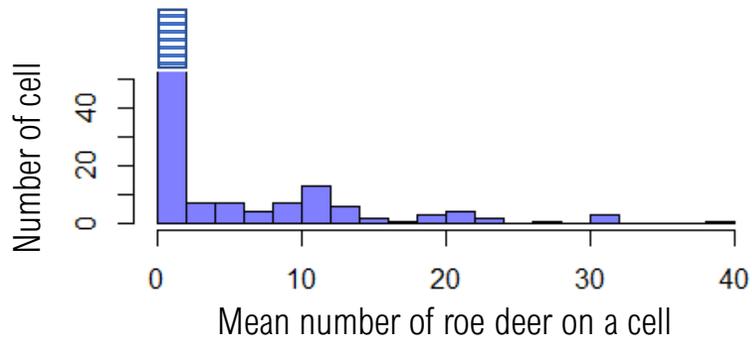
Sources: calculations from IGN data of the BDTPO

# Modelling roadkill : first results

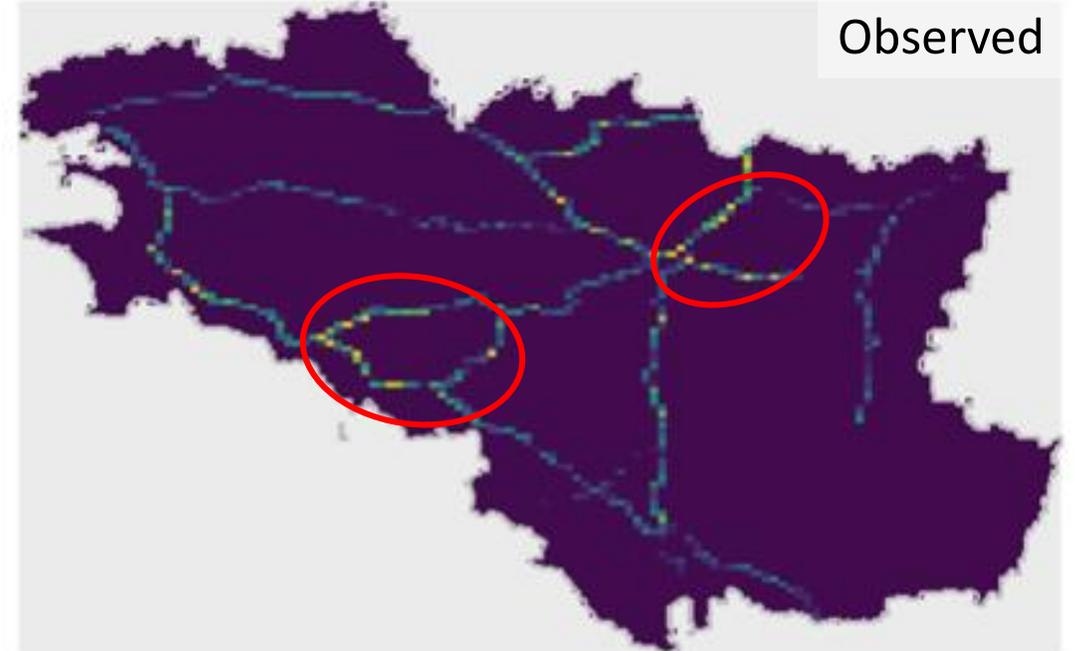


Number of roe deer roadkills simulated on 24km<sup>2</sup>

10 20 30

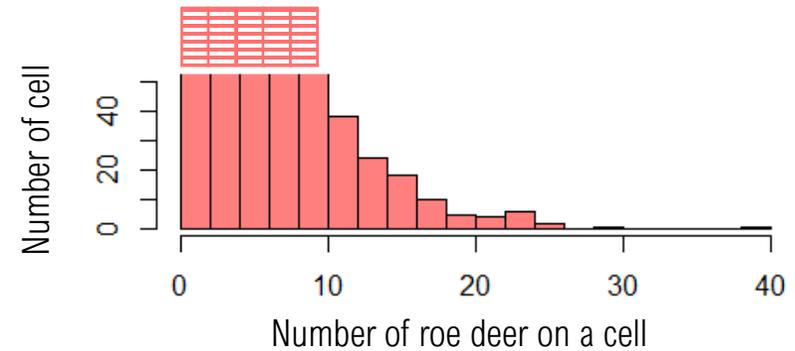


~



Number of roe deer roadkills observed on 24 km<sup>2</sup>

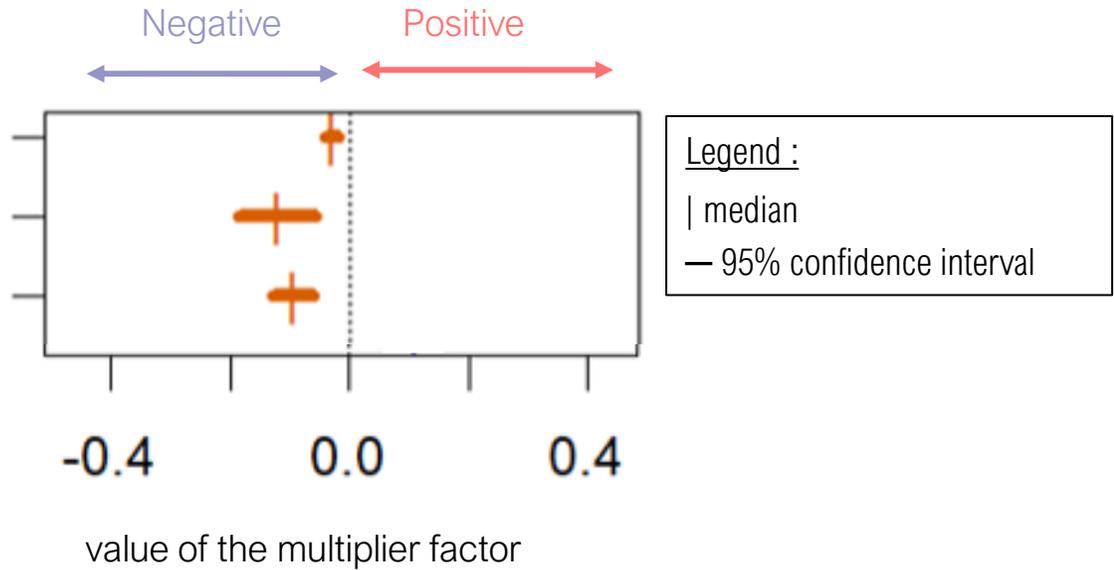
0 10 20 30



# Modelling roadkill : first results

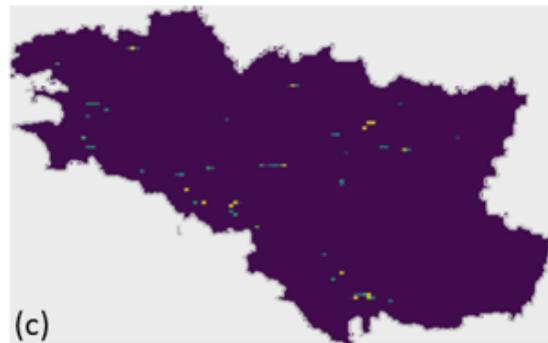
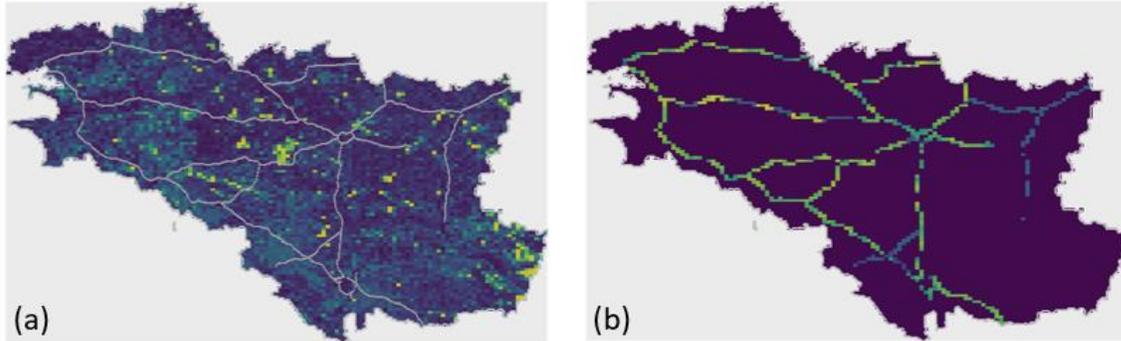


- Traffic
- Distance to nearest vegetation
- Distance to nearest waterway



Medium speed effect < High speed effect

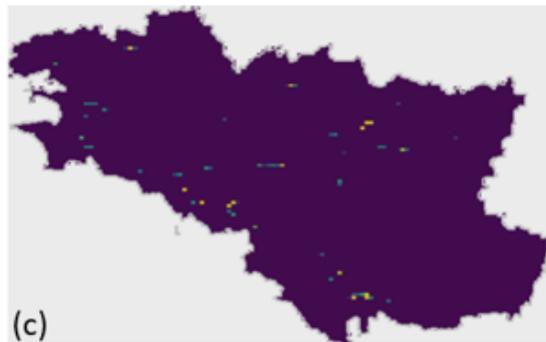
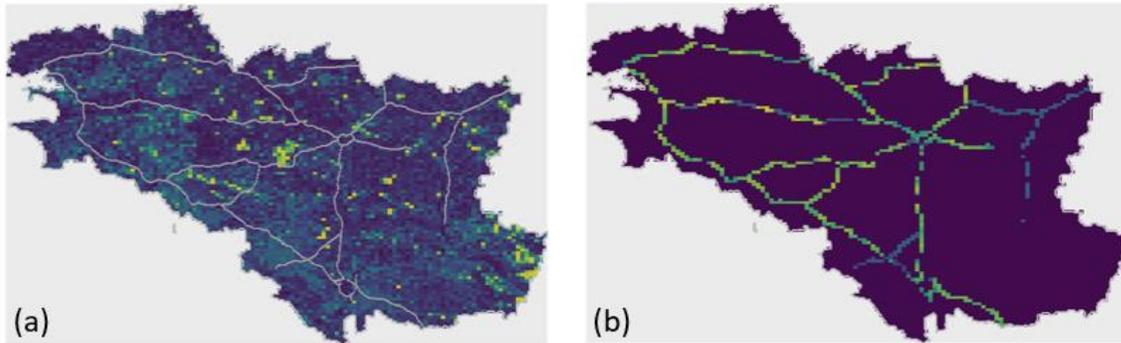
# Conclusion



## Achievements

- Provide a regional road risk map considering the species distribution in addition with the hazard.
- Build a framework for dealing with opportunistic data

# Conclusion



## Achievements

- Provide a **regional road risk map** considering the species distribution in addition with the hazard.
- Build a framework for **dealing with opportunistic data**

## Limits

- Single-species model
- Common species only
- No data proofing for now

## Prospects

- Improve the accuracy of the model
- Integrate other data sources in model
- Validate the model with protocolled data

## SCRIPT colloque sfeGFO\_conservationBiology\_spatialModelRoadkill

Hello to all of you, I will talk about 'Joint spatial modelling of wildlife casualties along major roads of Brittany'. This work was conducted in the COCPITT project which is an exploratory project funded by the ITTECOP program.

In a year, more than 5000 animal carcasses were recorded on Brittany's major roads, **an average of 3.5 per kilometer**. These events are highly problematic as they involve both conservation and safety issues. With the landscape fragmentation by roads, the wildlife roadkill caused is a threat to the integrity of the species by reducing the population size and genetic diversity, which can lead to local extinction of species. These accidents are also a source of material damage and and sometimes of bodily injury to road users.

The combination of these two issues makes it necessary to mitigate collisions. This is the target of the mitigation measures such as the wildlife underpass in this picture. Local authorities are looking for decision support tools at regional scale such as wildlife roadkill spatial prediction.

To access the collision risk, we modelled it in the framework of a risk analysis. To do that, we model the exposure which corresponds to where the animal is along the road. This allows us to take into account the fluctuation of animal density on the territory. The second component of the risk analysis is the hazard related to the road which will depend on its intrinsic characteristics.

We fit this model from two data sources in a joint model. The first data source considers the wildlife roadkill points on the road and informs the exposure and mainly hazard module. To accurately estimate exposure and separate its effect from that of the hazard, we will use the second data source that considers living wildlife observation points. The estimation of the parameters of this model is performed in a Bayesian way using the INLA-SPDE framework. For the next part of the presentation, I will present the model on the roe deer building and its preliminary results.

Now focus on the exposure. So let's first look at the data used to estimate it. Here the map shows live deer sightings in the study area. These points come from opportunistic data reported by amateur naturalists. They are referenced in the Visionature database which is coordinated by local naturalist associations. Those data are numerous and available at low cost for large

areas and time ranges and represent a true advantage. However, they are non-protocol and their distribution is dependent on the distribution and behaviors of the observers.

Therefore, we integrate a module on sampling effort to model these data. Once the sampling bias is isolated, the exposure, i.e. the intensity of occurrence, is modeled using a niche model. The predictors of this exposure are therefore the characterization of the habitat from landcover layers.

We observe the following results from the live roe deer modeling; dissimilar occurrence intensity patterns between model predictions (left) and observations (right). We hypothesize that this adjustment deficit is related to a difficulty in integrating the sampling effort. Indeed, in the histograms representing the average intensity per 24 km<sup>2</sup> grid cell, we see that the predicted intensities are all between 0 and 1, while the observations can reach very high values (more than 200 deer). The very high values of observations are mainly due to a strong sampling effort. Also, the predictors of deer presence are very consistent with its ecology. This consistency allows the roadkill risk to be properly informed about exposure.

Next we need to determine the second component of risk: the hazard of the road. For this we have semi-potocollated data on wildlife roadkill. On this map you can see the deer carcasses detected by road patrollers for the past 7 years. These patrols cover the entire region daily. But patrollers are not naturalists and carcass detection is a subsidiary task done at high speed. These last points can be problematic for the detection of small species but do not concern the roe deer since it is a common large species.

With these data we can adjust mainly the road hazard. The road hazard is predicted from the characteristics of a road sections like traffic, speed and by the proximity to a potential crossing point (vegetation or waterway).

We observe the following results from the roadkill modeling; some of the more deadly areas in the observations are partially found in the predictions, as you can see in the red circles. But the predictions detect too little the medium mortality areas (between 0 and 10 dead deer). We see this result in the histograms below. These bars with the hatched top mean that we find more than 40 cells with the corresponding number of deaths. For the observations (right histogram), we see that there are always more than 40 cells for each number of dead deer between 0 and 10. We do not find this high cell number of medium values in the predictions.

Vandroux Maëliiss

As was the case for exposure, the effects of the road hazard predictors are consistent with the literature. For example, high speeds are more dangerous, and traffic does not necessarily have a positive effect, because high traffic intensity can lead to road avoidance by animals.

To conclude, here is the map representation of our risk analysis. We were able to obtain a map of regional deer occurrence intensity from opportunistic data by isolating the sampling bias. From the joint modeling of exposure and hazard, we obtained a road risk map informed by both species distribution and road hazard.

This preliminary model has some limitations. First, it is a single-species model and the test species is a common species. But the validation of the model misses the robustness which brings the proofing of the data.

Currently, work is ongoing to find methodological solutions to improve the modeling accuracy of opportunistic data. But one of the main prospects is to think about how to confront protocol data to the model. This would allow to evaluate the predictive quality of the model.

Finally this approach opens the possibility of predicting ecological processes of conservation concern over large areas using multiple data sources even of variable quality.