

Cartes de densité de collisions faune-véhicules en France métropolitaine

basées sur des données semi-protocolées et non-protocolées



Agir pour
la biodiversité



Annaëlle Bénard (LPO AuRA, CNRS LEHNA)
Thierry Lengagne (CNRS LEHNA)

Avec le soutien de :



Direction Générale des Infrastructures, des Transports et des Mobilités



DREAL Auvergne-Rhône-Alpes
Direction régionale de l'environnement, de l'aménagement et du logement



Agir pour
la biodiversité

Pour citer ce travail :

LPO Auvergne-Rhône-Alpes - CNRS LEHNA (2025). Bénard A. & Lengagne T., *Cartes de densité de collisions faune-véhicules en France métropolitaine basées sur des données semi-protocolées et non-protocolées.*

Comptages routiers de carcasses	1
Prise en compte des biais	1
Contenu du livrable	2
Densité de collisions	3
Analyse de points noirs	4
Qu'est ce qu'un point noir ?	4
Application de la méthode aux données non-protocolées	4
Processus de correction : explications détaillées	7
Données opportunistes participatives Faune France (LPO)	7
Données des agents des DIR	9

Comptages routiers de carcasses

Les collisions faune-véhicules sont principalement étudiées au travers de comptages d'animaux écrasés visibles sur les routes. Cependant, un animal écrasé ne reste pas indéfiniment sur la route ; ce délai de disparition, couplé à la fréquence de passage des observateurs, affecte donc le nombre d'écrasements présents lors du comptage¹. Enfin, tous les écrasements présents sur la route ne sont pas systématiquement vus, en particulier pour les petites espèces². Ainsi, la **densité d'observations de carcasses** par kilomètre de route sur un réseau ne reflète pas directement la **densité réelle de collisions**. Pour obtenir une estimation fiable et standardisée à partir de données de comptage, il est nécessaire de connaître *a minima* les informations suivantes :

- La fréquence de passage des observateurs ;
- Le taux de détection spécifique à chaque espèce, basé sur les performances des observateurs ;
- Le temps moyen de disparition des carcasses dans la zone d'étude, calculé pour les espèces concernées.

Le relevé doit également être exhaustif et inclure toutes les espèces observées (dans le cas d'une étude portant sur un groupe d'espèces). Idéalement, ces relevés se réalisent à pied ou à vélo pour des portions courtes, ou en voiture à vitesse réduite avec un binôme composé d'un conducteur et d'un observateur, afin de détecter le maximum de carcasses. En disposant de ces informations, il devient possible d'estimer le nombre réel d'animaux écrasés par kilomètre de route. Avec cette méthode, il reste pour le moment impossible de comptabiliser les cas où l'animal s'éloigne de la route, reste coincé dans la carrosserie du véhicule ou est projeté loin de la chaussée après la collision, faute de connaissances à ce sujet³.

Ces protocoles, bien qu'essentiels, sont contraignants. En conséquence, les données disponibles en France ne respectent pas toujours ces exigences, ce qui nécessite des estimations pour corriger certains biais lorsqu'elles sont possibles, ainsi que l'acceptation des biais qui ne peuvent être estimés. En particulier, deux principales sources de données sont utilisées : les données issues des sciences participatives (comme celles collectées par les contributeurs du projet Faune France) et les relevés semi-protocolés réalisés lors de patrouilles régulières, notamment par les agents des Directions Interdépartementales des Routes (DIR). Afin d'exploiter ces données, il est indispensable de quantifier les éléments mentionnés précédemment (comme le taux de détection ou le délai de disparition), en s'appuyant sur des informations externes, des protocoles expérimentaux ou des modélisations.

Prise en compte des biais

La méthode proposée ici permet de corriger partiellement les biais qui induisent des différences entre les signalements et les collisions :

- Différences de visibilité selon les espèces ;
- Fréquence de passage des observateurs sur les routes.

Cependant, nous ne corrigons pas par :

- Le temps de persistance des carcasses, car les données expérimentales disponibles⁴ n'ont pas mis en évidence de structure spatiale (la persistance est la même sur toutes les routes au regard des informations disponibles) ;
- L'impact de la vitesse de déplacement des observateurs sur leur capacité à détecter les écrasements, ce facteur n'ayant pas été retenu dans nos expériences².

Néanmoins, la littérature scientifique indique que ces facteurs peuvent jouer un rôle dans le processus de relevé des carcasses^{1,5}, bien que cela n'ait pas été démontré dans ce travail.

¹Santos, 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(9). <https://doi.org/10.1371/journal.pone.0025383>

²Benard, A. (2023). *Road ecology: toward a predictive model of wildlife-vehicle collisions*, Chapitre 3 HAL: tel-04776552

³Román, J., Rodríguez, C., García-Rodríguez, A., Diez-Virto, I., Gutiérrez-Expósito, C., Jubete, F., Paniw, M., Clavero, M., Revilla, E., D'Amico, M. (2024). Beyond crippling bias: Carcass-location bias in roadkill studies. *Conservation Science and Practice*, 6(4), e13103.<https://doi.org/10.1111/csp2.13103>

⁴Benard, A. (2023). *Road ecology: toward a predictive model of wildlife-vehicle collisions*, Chapitre 2 HAL: tel-04776552

⁵Collinson, W. J., Parker, D. M., Bernard, R. T. F., Reilly, B. K., Davies-Mostert, H. T. (2014). Wildlife road

Contenu du livrable

Les données DIR et Faune France sont traitées à part, car leur hétérogénéité ne permet pas de créer des cartes conjointes. Ce livrable contient des cartes au format **geopackage**, basées sur la BD CARTO® IGN 2021. En parallèle, les cartes sont accessibles via le SIG public de la LPO Auvergne-Rhône-Alpes : <https://sig.lpo-aura.org/index.php>.

Les cartes disponibles incluent les espèces à enjeu de conservation (PNA en cours, liste rouge,...) et les groupes d'espèces suivants :

- Mammifères : Blaireau européen, Castor d'Eurasie, Cervidés, Chauve-souris, Chat forestier, Hérisson d'Europe, Lagomorphes, Lynx boréal, Loutre d'Europe, Mustélidés, Renard roux, Rongeurs, Sanglier.
- Oiseaux : Canards, Chevêche d'Athéna, Colombidés, Corvidés, Échassiers, Laridés, Limicoles, Martin-pêcheur d'Europe, Œdicnème criard, Oiseaux d'eau (autres), Passereaux, Pie-grièche, Rapaces diurnes, Rapaces nocturnes.
- Amphibiens et Reptiles : Anoures, Cistude d'Europe, Crapaud vert, Lézard, Lézard ocellé, Salamandre, Serpents, Sonneur à ventre jaune, Triton.

Les espèces sont également regroupées selon les types d'aménagements routiers à envisager, tels que :

- Passage 'toute faune'.
- Passage 'petite faune' hydraulique avec banquette ou encorbellement.
- Passage 'petite faune' agricole ou forestier.
- Écoduc (incluant passage pour amphibiens).
- Écuroduc et Chiropteroduc.
- Gestion de la végétation.

Ces groupes ont été établis avec l'expertise du CEREMA (E. Guinard et J-F. Bretaud).

Densité de collisions

Pour chaque tronçon de route, le fichier comprend la densité de collisions (**densité de signalement corrigée pour la détection des écrasements et la fréquence des visites**, voir la section *Explications détaillées*) correspondant à chaque année. Il est également possible d'afficher la densité pour toutes les années confondues (données Faune France uniquement, voir *Données des agents des DIR*)

Détails techniques

Chaque carte s'appelle `group_hm.gpkg`. Elle contient les colonnes suivantes : `transect_id`, `class_adm` (provenant de la cartographie IGN), région, longueur du tronçon (km), et `dc_année`, qui correspond à la **Densité Corrigée** et normalisée de collisions. Cette densité est normalisée, elle est donc sans dimension entre 0 et 1, où des valeurs proches de 1 indiquent une densité de collisions plus élevée. Les valeurs ne sont pas directement comparables entre plusieurs cartes (par exemple, une densité de 0,5 sur une carte de collisions Blaireau européen n'équivaut pas à une densité de 0,5 sur la carte des Rapaces nocturnes).



Exemple de la carte de densité toute espèces pour les données LPO, visible en ligne. Pour alléger la représentation, seules les routes avec une densité corrigée de collisions non nulle sont affichées.

Analyse de points noirs

Qu'est ce qu'un point noir ?

En outre, les résultats d'une analyse regroupant toutes les années à l'aide de l'outil KDE+⁶ sont présentés pour chacune des cartes proposées ci-dessus. Les points noirs d'écrasement routier, sensu KDE+, sont définis comme des agrégations spatiales d'écrasements où la densité est trop élevée pour résulter d'une répartition aléatoire, selon un test statistique d'hypothèse avec un seuil $\alpha = 0.05$. Ces points noirs indiquent des **lieux précis (quelques centaines de mètres) où le nombre d'écrasement de la faune est important par rapport au reste de la section de route** (section : portion entre deux intersections). Leur présence résulte d'une interaction complexe entre des facteurs locaux spécifiques (comme la proximité d'une forêt, la présence de végétation linéaire ou de cours d'eau, ou encore des caractéristiques routières comme des talus), et non d'une simple répartition uniforme ou aléatoire des collisions.

Par conséquent, les points noirs constituent des outils essentiels pour orienter les efforts de prévention et de gestion des risques. Un point noir de collisions ne se trouve pas nécessairement sur un tronçon de route où la densité de collisions est élevée ; à l'inverse, un tronçon présentant une forte densité de collisions ne contient pas forcément de point noir. En effet, certaines portions de route peuvent enregistrer un faible nombre de collisions, mais fortement concentrées sur quelques mètres, tandis que d'autres tronçons peuvent présenter un grand nombre de collisions réparties uniformément, donc sans point noir. Dans ce second cas, les collisions sont plus difficiles à cibler avec des aménagements précis tels que des passages à faune. Cela n'exclut pas la nécessité de réfléchir à des mesures pour atténuer l'impact de ces zones étendues à forte densité de collisions. Cependant, **lorsqu'un point noir est identifié sur un tronçon présentant également une densité élevée, il peut constituer une priorité stratégique.**

Application de la méthode aux données non-protocolées

Les biais cités précédemment ne s'appliquent pas de la même façon dans cette analyse. Nous travaillons à l'échelle d'un tronçon de route entre deux intersections, où le nombre d'observateurs est, par définition, constant (aucun nouveau véhicule ne peut entrer ni sortir de cette portion). Par conséquent, l'analyse KDE+ n'est pas biaisée par la fréquence de passage sur les routes. De plus, la prise en compte des changements de conditions de conduite (notamment les limitations de vitesse) permet d'homogénéiser encore davantage les probabilités de détection des carcasses. Cependant, comme le soulignent les auteurs de la méthode⁷, une hétérogénéité peut subsister, en particulier en raison des différences de détection, de persistance et de volonté de signalement entre les espèces (pour les analyses portant sur des groupes d'espèces), et des variations temporelles dans l'effort d'échantillonnage. Ces facteurs peuvent engendrer de faibles taux de faux positifs (points noirs identifiés à tort) ou des faux négatifs (absence de points noirs malgré leur présence réelle). De plus, cette méthode n'est capable d'identifier des points noirs que si les données disponibles sont suffisantes. Les espèces difficiles à recenser, notamment les très petites espèces qui sont particulièrement difficiles à repérer, peuvent avoir des points noirs de collisions qui passent inaperçus en raison d'un nombre insuffisant de données.

Ainsi, bien que la méthode soit robuste pour une application sur des données non ou semi-protocolées, elle n'est pas infaillible et doit être utilisée comme un **outil d'aide à l'identification des tronçons de route dangereux, plutôt que comme un guide absolu pour la mise en place d'aménagements.**

⁶Bíl, 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. <https://doi.org/10.1016/j.aap.2013.03.003>

⁷Bíl, M., Andrášik, R. (2020). The effect of wildlife carcass underreporting on KDE+ hotspots identification and importance. *Journal of Environmental Management*, 275, 111254. <https://doi.org/10.1016/j.jenvman.2020.111254>

Recommendations du CEREMA

La détection d'un point noir sur le réseau routier devra être suivie d'un diagnostic fin du secteur collisiogène (dans un rayon de 300m pour la petite faune, jusqu'à 3km pour la grande faune) avec notamment pour objectifs :

1. Identifier les ouvrages d'arts existants non dédiés et la possibilité de leur aménagement (veiller à évaluer l'utilisation de l'existant par la faune pour alimenter le suivi d'efficacité post-aménagement).
2. Identifier les équipements de la route sur lesquels s'appuyer pour diriger les espèces vers un passage sécurisé (ex : complément d'engrillagement).
3. Évaluer les possibilités d'aménagements neufs (ex : écoduc).

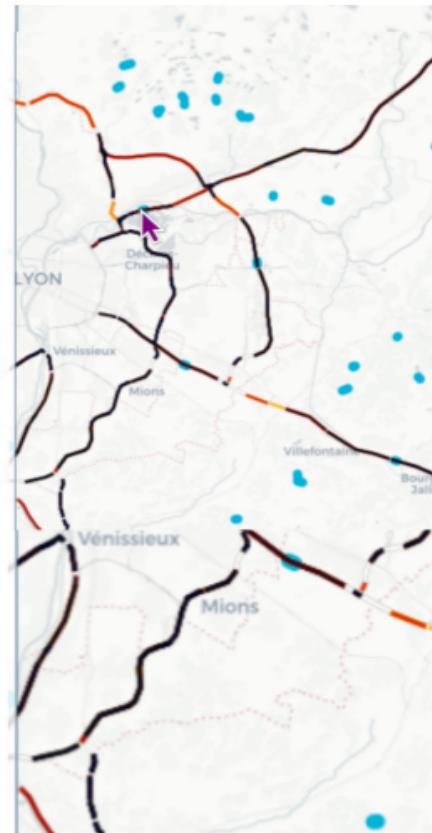
Pour le dimensionnement des ouvrages, cf. : <https://www.cerema.fr/fr/centre-ressources/boutique/passages-faune>.

Détails techniques

Uniquement pour les espèces ou groupes d'espèce pour lesquelles des points noirs ont été détectés.

Chaque carte est nommée *groupe_kde.gpkg*, et contient les points noirs retenus (Sstr = 1, Nombre de points par cluster ≥ 4). Chaque point noir contient les informations suivantes : NPts_clus (nombre de points du cluster) et Strength⁸, et est accompagné

- d'un diagramme de répartition des espèces concernées (proportions corrigées par le taux de détection des espèces depuis une voiture), nommé "*ID du cluster_pie.png*" ;
- d'un diagramme temporel indiquant les années pour lesquelles des données du point noir ont été signalées, nommé "*ID du cluster_temps.png*".



Exemple de carte en ligne. Les points noirs identifiés sont superposés à la carte de densité. Lorsque la carte est dézoomée, seules les densités corrigées pour les routes principales s'affichent, mais tous les points noirs restent visibles. Un clic sur le point noir affiche la répartition des espèces ainsi que l'évolution des signalements dans le temps.

⁸Se référer à la documentation technique de l'outil KDE+ pour plus de détails [<https://www.kdeplus.cz/en/faq>]

Processus de correction : explications détaillées

Données opportunistes participatives Faune France (LPO)

Les étapes de correction sont les suivantes :

1. Suppression des observations non rattachées à une route, c'est-à-dire des signalements sans précision GPS (<2% des signalements totaux).
2. Suppression des doublons. Les doublons sont définis comme des signalements de la même espèce faits par deux contributeurs différents, situés à moins de 200 mètres et reportés à moins de 4 jours d'intervalle (95% quantile de persistance des espèces⁹). Une exception est faite pour les amphibiens, souvent écrasés en groupe.

Ce jeu de données ne repose sur aucun un protocole de comptage, il nous manque en particulier des informations sur la fréquence de passage des observateurs. Pour contourner ce biais, nous utilisons un modèle statistique, basé sur les déplacements de plusieurs dizaines de contributeurs Faune dans la région Auvergne-Rhône-Alpes, relevés au sein de l'expérience *transmorta* menée via l'application Naturalist. Ce modèle permet de prédire les routes du réseau français les mieux surveillées par les contributeurs.

En complément, une correction est appliquée grâce aux taux de détection en fonction de la taille et du taxon des espèces, basée sur une expérience menée auprès des contributeurs Faune¹⁰. Ces deux corrections produisent une **densité relative de collisions par tronçon de route** (fondée sur la cartographie BD CARTO® IGN, 2021), exprimée sans dimension entre 0 (faible densité) et 1 (forte densité).

Prédiction du niveau de surveillance des routes

Pour prédire le niveau de surveillance de chaque route du réseau par les contributeurs Faune, nous avons testé 9 modèles mécanistes et 2 types de modèles statistiques. La performance de ces modèles est évaluée par comparaison à un jeu de données de validation, ci-après *transmorta*, récolté grâce à la participation d'une grande partie des salariés LPO AuRA et de quelques adhérents, qui ont effectué des comptages routiers standardisés (enregistrement des observations d'écrasements et des trajets effectués). Au total, 250 trajets totalisant près de 9000 km ont été enregistrés dans la région sur une période de plusieurs semaines à l'hiver 2024.

Les modèles sont évalués sur plusieurs métriques :

- Corrélation de Pearson : correspondance entre le nombre de visites prédites et constatées sur chaque tronçon de route.
- Recall : capacité à détecter tous les tronçons effectivement surveillés.
- Precision : capacité du modèle à éviter de classer à tort une route comme surveillée.
- Score F1 : mesure équilibrée entre le Recall et la Sensitivity.
- R² : degré d'adéquation entre les prédictions et les données réelles.

Ces métriques permettent d'évaluer la capacité des modèles à classer les routes selon leur niveau de surveillance, à minimiser les taux de faux positifs et de faux négatifs, et à prédire précisément le nombre de visites par tronçon de route. Les modèles statistiques sont également optimisés par validation croisée, une technique consistant à diviser les données en plusieurs sous-ensembles pour réduire les risques de surentrainement. Les modèles mécanistes incluent :

- un modèle nul, attribuant un nombre de visites aléatoires sur chaque tronçon de route selon une loi de Poisson paramétrée sur les données *transmorta*.
- deux modèles basés uniquement sur le trafic routier : nombre de visites proportionnel au volume de trafic (loi uniforme), et attribution des visites selon une loi de Poisson proportionnelle au volume de trafic.
- des modèles basés uniquement sur l'activité des contributeurs dans Faune, qui se concentrent sur le nombre de visites inférées à partir des observations d'écrasements et de l'activité de report d'animaux

⁹Benard, A. (2023). *Road ecology: toward a predictive model of wildlife-vehicle collisions*, Chapitre 2 HAL: tel-04776552

¹⁰Benard, A. (2023). *Road ecology: toward a predictive model of wildlife-vehicle collisions*, Chapitre 3 HAL: tel-04776552

vivants sur le territoire. Par exemple, une observation d'animal *vivant* peut conférer une visite aux routes situées dans un rayon de 100 m, 500 m ou 1000 m, ou aux 3 ou 5 routes les plus proches, avec ou sans pondération pour attribuer davantage de visites aux routes les plus proches.

Deux types de modèles statistiques sont testés, tous deux entraînés sur les données *transmorta*. Le premier type est un modèle linéaire généralisé basé sur une loi de Poisson *zero-inflated*. Le second type est un modèle de type *Random Forest*. Les modèles utilisent des prédicteurs tels que la classe administrative de la route (nationale, départementale...), la vocation de la route (liaison locale, régionale...), la présence ou non de péages, le nombre de voies, la densité de population locale (données INSEE) et l'activité des contributeurs, calculée à partir des observations d'écrasements ou d'espèces vivantes. Dans le cas du modèle linéaire, l'absence de multicolinéarité entre les prédicteurs a été vérifiée. La décision d'inclure certains prédicteurs, comme les caractéristiques des routes et la densité de population, proviennent de travaux précédents sur la modélisation du volume de trafic routier annuel.

Les modèles mécanistes montrent de faibles performances sur toutes les métriques évaluées. En particulier, les modèles basés uniquement sur le volume de trafic n'expliquent pas correctement les visites sur les routes, révélant que les déplacements des contributeurs Faune ne sont pas directement corrélés à ceux des autres utilisateurs de la route. Le modèle retenu est un modèle de type *Random Forest*, intégrant des prédicteurs tels que l'activité des contributeurs concernant les espèces vivantes (transformée en variable continue grâce à une fonction KDE), le signalement des écrasements, le nombre de voies, la classe administrative, la présence de péages et la densité de population (terme quadratique). Il montre une corrélation de Pearson de 0.85 (moyenne calculée sur une validation croisée à 10 plis), un score F1 moyen de 0.75, équilibré en Recall et Precision. En revanche, le R^2 de 0.3 met en évidence les limites du modèle à prédire avec précision le nombre absolu de visites sur chaque route. Ce résultat justifie en partie la décision d'utiliser des densités corrigées sans unité, car le modèle est mieux adapté à estimer un niveau relatif de surveillance des routes par rapport au reste du réseau, plutôt qu'à fournir des prévisions quantitatives exactes.

Ce modèle utilise 200 arbres, un nombre sélectionné en visualisant les variations du *Out Of Bag MSE*, qui représente une mesure de la capacité du modèle à bien généraliser sur des données non vues. Le nombre de prédicteurs à considérer pour chaque division est fixé à 5. Le modèle prédit notamment un nombre élevé de visites pour les routes majeures telles que les autoroutes et les nationales, ainsi que pour les routes situées dans des zones de forte biodiversité, comme la Dombes dans le département de l'Ain. Bien sûr, il ne s'agit que d'une modélisation : bien que performant au regard des métriques utilisées, le modèle peut sous ou sur-estimer le taux de surveillance des routes du réseau routier français. En particulier, le modèle est entraîné sur le réseau routier de la région Auvergne-Rhône-Alpes avec les contributeurs de projet les plus actifs, mais il est probablement moins performant quand il s'agit de généraliser aux autres régions.

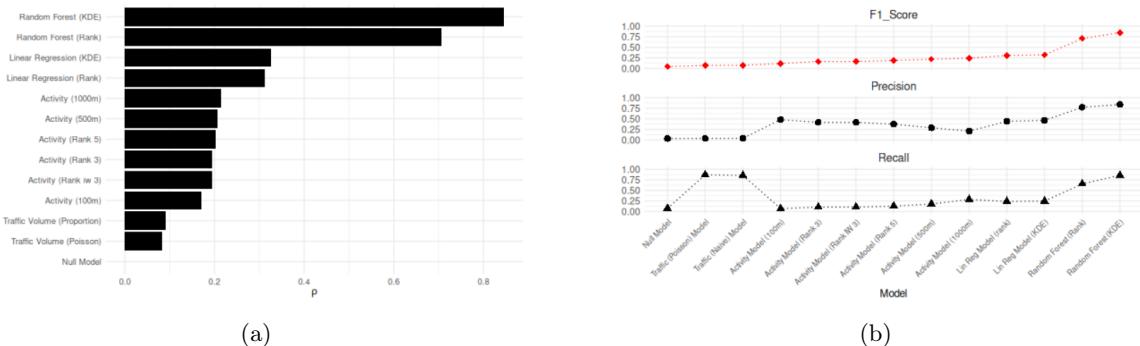


Figure 1: Performance des modèles utilisés pour prédire le niveau de surveillance des routes par les contributeurs de Faune. (a) Corrélation entre le niveau de surveillance mesuré et prédit, (b) Métriques de performance.

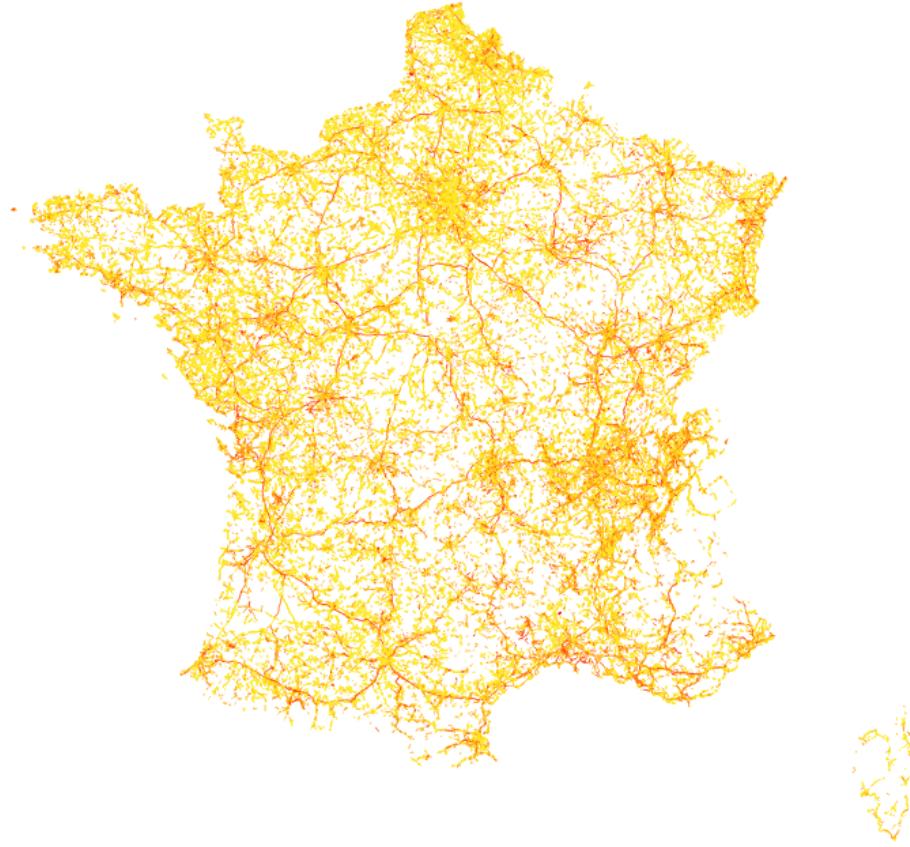


Figure 2: Carte de la prédiction du niveau de surveillance en 2023 (jaune = peu surveillé, rouge = fréquemment surveillé) d’après le modèle retenu.

Données des agents des DIR

Ce jeu de données n'est pas entièrement protocolé. Les agents reportent les espèces écrasées rencontrées lors de leurs patrouilles, mais nous ignorons leur fréquence de passage réelle sur les routes. Nous disposons néanmoins des informations suivantes :

- Au niveau national, un cadrage théorique prévoit une fréquence de patrouille deux à trois fois supérieure lorsque le trafic dépasse 25 000 à 35 000 véhicules/jour (seuil retenu : 30 000 véhicules/jour). Les patrouilles ajoutées ponctuellement ne sont pas prises en compte faute de données détaillées.
- La DIR Nord-Ouest fournit ses Plans de Déclinaisons de la Viabilité, détaillant les fréquences de patrouille par route.

Ces données permettent de créer une carte des fréquences théoriques de passage sur le réseau national. Cependant, ces éléments théoriques ne reflètent pas nécessairement la réalité. Idéalement, le nombre réel de patrouilles par unité de temps (mois, année) pour chaque route serait nécessaire. En effet, une différence entre le nombre de passage théorique et la fréquence réelle des patrouilles mènera fatalement à une **sur ou sous-estimation de la densité de collisions**, selon que le réseau est moins surveillé, ou plus surveillé que prévu.

Nous appliquons également une correction des taux de détection des espèces, basée sur les différences relatives observées dans nos expériences⁸. Bien que ces expériences n'aient pas impliqué directement les agents DIR, ce qui nous empêcherait d'accéder aux taux de détection absolu, nous supposons que les différences relatives entre les espèces (*e.g.*, un chevreuil est vu deux fois plus souvent qu'un hérisson) sont applicables.

Les biais éventuels suivants ne peuvent être corrigés faute d'informations :

- Différences de volonté de signalement selon les espèces ;
- Proportion de patrouilles sans activité de relevé ;
- Signalement répété d'un même animal écrasé (doublon).

En effet, certaines DIR semblent plus assidues dans le report exhaustif des espèces, et le relevé des données n'est pas uniforme sur toute la période étudiée. Par exemple :

- De 2011 à 2013, seule la DIR Est remonte des données ;
- De 2014 à 2015, seules les DIR Est et Ouest semblent actives ;
- De 2016 à 2017, seules les DIR Est, Ouest, Atlantique et Nord semblent actives.

Ces disparités empêchent de calculer une densité de collisions compilée pour la période 2011-2022 (*dc_all*), qui serait biaisée en faveur des DIR actives sur une plus longue durée. En revanche, les densités de collisions année par année sont accessibles. De même, la carte de densité de collisions pour certaines espèces semble biaisée en faveur de certaines régions, indiquant que certaines DIR sont probablement plus assidues dans le signalement de ces espèces que les autres. Les cartes de densité de collisions sont exprimées sous forme de valeurs comprises entre 0 et 1, sans unité, en grande partie en raison de l'absence de données sur la performance de détection absolue des agents DIR. Une évaluation comparable à celle réalisée avec les contributeurs Faune France serait nécessaire pour combler cette lacune.