**Spatial analysis of the wildlife roadkill risk at a regional scale.**

7000 mots au total de la page de titre à la dernière figure (hors annexes). Interligne double, manuscrit anonymisé, format d’abstract imposé

**Authors**

Papaïx, Julien1\*

Barnagaud, Jean-Yves2\* ORCID 0000-0002-8438-7809

Vandroux, Maëliss2

[1 à 3 personnes de la LPO et Bretagne Vivante à décider – important pour le journal cible]

Bretaud, Jean-François

Morand, Alain

\* shared first authorship

**Corresponding author**

Jean-Yves Barnagaud – jean-yves.barnagaud@ephe.psl.eu

**Affiliations**

Ajoutez vos affiliations SVP – NB : Maëliss doit être affiliée conjointement au CEREMA et au CEFE.

1: INRAE…

2: CEFE, Univ Montpellier, CNRS, EPHE-PSL University, IRD, 1919 route de Mende, 34293 Montpellier, France.

**Acknowledgements**

This study was funded by the ITTECOP program under exploratory grant n°XXX – COCPITT, and received support of the Fondation pour la Recherche sur la Biodiversité through the CESAB/Centre for the Synthesis and Analysis of Biodiversity (FRB/MTE/OFB: project ACOUCENE). The authors thank ITTECOP’s deputy director, Yannick Autret, for his support throughout the project, Laurent Couzi (LPO France) for establishing excellent relationships with local delegates of the LPO and Bretagne Vivante, and Olivier Gimenez (CNRS) for triggering initial contacts between the authors. This study is dedicated to the thousands of naturalist enthusiast who gather opportunistic wildlife data through the VisioNature portal and to the members of the GDR Ecostat and Réseau CISSTAT who engage into science-stakeholders collaborations through data analysis. [ajoutez qui vous voulez : DIR, etc etc].

**Author contributions**

JP : wrote the model and performed the analyses, led the writing.

JYB : conceived the initial idea and led the project, contributed to theoretical and methodological developments, led the writing.

AM : conceived the initial idea, contributed to applied developments, participated to the writing.

JF : contributed to applied developments, gathered and explored spatial data, participated to the writing.

MV : prepared and explored data, implemented the initial versions of the model and performed the interpretation of initial results, participated to the writing.

**Statement on inclusion**

Our study was co-constructed from the beginning between academic researchers in community ecology and spatial statistics, engineers from regional agencies and two non-governmental associations, Bretagne Vivante and the Ligue pour la Protection des Oiseaux – Pays de Loire, with the support of the national Ligue pour la Protection des Oiseaux and the French Ministry in charge of Environment. Non-academic regional stakeholders directly participated to the conception of the model and analyses by providing their local expertise on data and results and are included in the authors list. Regular meetings and written contacts allowed a constant flow of exchange between researchers and regional stakeholders through the student in charge of the project (MV, included in the authors list).

**Data availability statement**

Upon publication, the data and scripts allowing the replication of the results presented in this study will be deposited on the French government’s public research archive repository:

<https://www.data.gouv.fr/fr/>

**Conflict of interest statement**

The authors declare no conflict of interest.

**Abstract** (350 wds – le nombre de mots sera ajusté une fois le contenu calé)

1. Spatial assessments of human imprints on ecosystem are needed to set relevant planning policies to preserve biodiversity over regional scales. However, many ecological processes are too elusive or sparsely distributed to be observed through protocoled sampling over large spatial extents with a reasonable effort and cost. An alternative is to rely on the large number of opportunistic data gathered by amateur naturalists and ecological practitioners, at the cost of inflation in model complexity and uncertainty in inference or prediction.
2. Here, we investigated the potential of opportunistic and semi-protocoled data to feed a regional-scale spatial model of the wildlife roadkill risk, a process of concern for road safety as well as for planning policies promoting ecological connectivity.
3. We modelled roadkill risk as the product of two point-pattern models describing respectively drivers’ exposure to a species’ occurrence and danger of collision, accounting for predictors of species’ distributions, proxies of road characteristics, sampling error and unexplained spatial autocorrelation.
4. We ran the model for seven mid-sized mammal species over a regional extent corresponding to Brittany, an administrative region of western France, focusing on the highest-traffic road both to avoid under-sampling and to be consistent with the priorities of regional road infrastructure policies.
5. [Results of the model 🡪 Julien / il faut qu’on decide ce qu’on met comme résultats ici]
6. We showed that biologically sensible predictive models can be built to inform regional policies on priority areas for the wildlife roadkill risk, even with non-protocoled data gathered opportunistically. Nevertheless, acquiring subsets of protocoled data to estimate the numerous sampling biases that impair detection both of exposure and danger is critically needed to increase the resolution of predictions. Furthermore, substantial effort should be put on data quality and homogeneity across the region to limit the need of data filtering and aggregation.
7. **Synthesis and applications.** Regional predictive maps relying on opportunistic and semi-protocoled data can serve as a template of interactions between field practitioners, policy makers and spatial modellers to jointly delineate priority areas for conservation actions and identify data weaknesses.

**French abstract [sera traduit une fois stabilisé]**

**Keywords**

Regional infrastructure, INLA-SDEP, opportunistic records, risk analysis, roadkills, spatial model

**Introduction**

Wildlife roadkills are a conspicuous expression of ecological threats associated with linear infrastructures, motivating research on risk assessment and mitigation strategies (Laurance and Balmford 2013, Van der Ree et al. 2015, Haddad 2015, Balčiauskas et al. 2025). Estimates of collisions with animals on European roads reach 194 million dead birds and 29 million mammals per year (Grilo et al. 2020), with substantial variation among taxonomic and functional groups (Rytwinski and Fahrig 2015, González-Suárez et al. 2018). With the generalized increase of road density, especially in densely-populated areas such as western Europe, these considerable figures are causing concern for public safety, demography of involved species, and the associated ecosystem processes (see e.g. Gkritza et al. 2010, Meijer et al. 2018). However, the balance between the development of transportation networks and the mitigation of disruptions in ecological connectivity caused by roads still requires macro-level assessments to be useful for the setting of green and blue infrastructures. Spatial models of the roadkill risk are therefore needed to incorporate this particular threat to wildlife into planning policies and envision spatially coherent responses over large scales (van der Ree et al. 2011, Grilo et al. 2025).

As a general pattern, roads and other linear infrastructures are viewed as physical disruptions of ecological connectivity that modify or prevent animal movements and gene flow, contribute to land artificialization and resource loss, and trigger direct mortality (Benitez-Lopez et al. 2010, Jackson and Fahrig 2011). These effects add up with habitat reduction and fragmentation within landscapes and are assumed to contribute substantially to the global depletion of some animal populations (Grilo et al. 2021). Quantifying accurately the impact of roads on wildlife is however challenging, as are their higher-level consequences for demographic rates and population densities. In particular, direct mortality of animals through roadkills only account for a fraction of the ecological threats posed by the ever-densifying network of roads in developed countries. As such, it probably has a limited impact on most common species’ population sizes beyond local inflations of the baseline mortality rate or disruption of dispersal corridors (Forman and Alexander 1998, Kang et al. 2016, González-Suárez et al. 2018). Nevertheless, roadkills trigger concern from public authorities in charge of transportation infrastructure because of the road safety issues involved by collisions with large animals. Furthermore, roadkills are conspicuous to the general public and relatively easier to monitor as compared with less direct impacts of roads on wildlife caused by connectivity disruptions, air or acoustic pollution. Mitigation actions have thus been engaged, but they appear heterogeneously effective and insufficiently evaluated (Rytwinski et al. 2016, Coulson and Bender 2019). They typically include fences and wildlife crossings, which are costly and sometimes technically challenging to set up. Their location along road networks thus need to be carefully planned (van der Grift et al. 2013, Polak et al. 2014, Ascensão et al. 2021), which requires spatial prediction of roadkill patterns over large spatial extents.

Building macro-scale models of wildlife roadkill patterns however remains a challenge, due to the complexity of underlying processes and the resulting difficulty of acquiring suitable data (Malo et al. 2004, van der Grift et al. 2013, Boyle et al. 2017). Roadkills are driven by multiscale, temporally variable process, which are further modulated by an intricated conjunction of landscape-level, habitat-level and road-specific factors. Anthropogenized landscapes such as urbanized areas, cultivation mosaics and orchards are notably related to an inflation of mortality along road for some species, but comparative analyses also reveal that the environmental determinants of roadkills depend on species, their traits, and seasons (Polak et al. 2014, Ha and Shilling 2018, Valerio et al. 2021). Similarly, although traffic, road type and density, and roadside characteristics are identified are key predictors, their effects are heterogeneous among regions and species (Bissonette and Kassar 2008, Benítez-López et al. 2010, Gunson et al. 2011, Kušta et al. 2017). Still, collision rates between large mammals and vehicles seem overall higher on national roads and motorways than on local networks (although this can also be due to undersampling of the latter), while traffic volume could have, at least for some species, a unimodal effect with reduced effect on the least and most frequented roads (Denneboom et al. 2024). As a general picture, spatial roadkill patterns thus appear to be explained by idiosyncratic, locally varying determinants that cannot be easily used for prediction with simple correlative models elaborated at coarse spatial resolutions and large extents.

One of the limitations impairing the identification of regional predictors may be that several processes have often been mixed into a single roadkill distribution pattern, often formed by a set of carcass locations (e.g. Malo et al. 2004, Santos et al. 2018, Pinto et al. 2023, Laube et al. 2023). This approach confounds the processes that set species’ geographical distributions and those that govern collisions between animals and vehicles, which need not to be the same. Studies separating these two processes are uncommon and usually dissociate them into an envelope model and a roadkill model without an explicit structure to allow for error propagation between them (Santos et al. 2013, Fabrizio et al. 2019, Russo et al. 2020, Frangini et al. 2022). This two-step process fails to recognize roadkill patterns as the stochastic realization of an interaction between two latent layers, drivers’ exposure to animals and probability of collision per-se (or danger).

A more explicit model can be derived from risk analysis, in which the intensity of the carcass point pattern at a given location *i*, a date *t* and for a species *s* () is modelled as , in which *E* is the drivers’ exposure to the species and *D* is the level of danger (Ahmed et al. 2021, Gurumurthy et al. 2022). Importantly, in this framework, *E* and *D* are two independent components of *R*, both allowed to have their own spatial structures and predictors, increasing biological realism with key implications for risk reduction strategies (Plante et al. 2019). For instance, a roadkill hotspot explained by high densities of a given species in a landscape (predicted by *E*) may be mitigated by fencing and building crossing structures, while high danger (predicted by *D*) would primarily require measures oriented towards vehicle drivers, traffic modulation or roadsides management.

A difficulty of roadkill patterns lies in that the spatial variance structure of the *E* and *D* layers differ, the former being a two-dimensional surface of presence intensity of live animals while the latter is organized along a one-dimensional linear network. Most studies have overcome this issue by downgrading the roadkill point process into a count of roadkills in a two-dimensional grid, using buffers around roads, or simply ignoring it (e.g. Grilo et al. 2015, Wright et al. 2020, Morelli et al. 2020, Martins et al. 2024). This way of doing permits to exploit usual two-dimensional models, at the cost of degraded resolution and suboptimum representation of the actual data generation process, possibly leading to predict roadkills out of the road network or to generate inadequate spatial variance structures [REFs – Julien si tu peux renforcer ici stp]. Conversely, the risk model framework distinguishes *E* and *D* into two independent models. So doing, *E* can be defined as a two-dimensional surface of probability of presence fitted with data on live animals through a patch-occupancy model (Altwegg and Nichols 2019), a point pattern models implemented *e.g.* by MaxEnt (Renner and Warton 2013) or any other envelope model (Guisan 2017). Because both layers are independent, *D* can be fitted through a one-dimensional spatial model, such as point pattern models for linear networks (Baddeley et al. 2015), allowing to constrain spatial error propagation pattern according to the road network.

[Julien : faire un point sur les solutions stats et pourquoi c’est plus compliqué que ça en a l’air – processus agrégé, structures 2D/réseau linéaire, propagation des erreurs entre les couches, etc, bref ce que tu penses être pertinent.]

A distinct issue is that the animal distribution and roadkill data used to fit *E* and *D* are affected by different survey biases, including search-location biases, uneven detection and unequal carcass persistence over time (Lima Santos et al. 2016, Román et al. 2024). While standardized protocols can be set up locally to account for these sources of uncertainty, they are out of reach at a regional extent. Most often, roadkill datasets usable at that scale are produced by road patrols or amateur naturalists with limited standardizations (Shilling et al. 2015, Swinnen et al. 2022, Grilo et al. 2025). These data are plagued with the usual sources of temporal and spatial biases associated with opportunistic records, including uneven sampling effort in space and time, biased search behaviour and limited replication (Mair and Ruete 2016, Tiago et al. 2017, Geurts et al. 2023). Occurrence records to evaluate *E* may even be more limiting since distribution data on many species are lacking proper spatial coverage and absence records (Laube et al. 2023). Species distribution models using opportunistic records, both for live animals or roadkills, often compensate these limitations through resampling or statistical procedure to account indirectly for sampling effort and missing absence data (Isaac et al. 2014, De Solan et al. 2019). However, when processed and interpreted with care, these messy data can provide reliable and useful insights on spatial roadkill patterns, especially over the large spatial extents and coarse resolutions required for regional planning policies (Isaac et al. 2014, Dobson et al. 2020, Shin et al. 2022).

This study originates from the French Ministry in charge of ecology’s interest for a regional understanding of the ecological impact of linear infrastructure networks. In particular, maps of the roadkill risk are requested by stakeholders to identify mortality hotspots along the French road network, encourage protocoled sampling in these areas, and implement targeted risk reduction actions. Thus, our aim was to evaluate the ability of opportunistic records gathered by volunteers and road safety patrols to produce a regional-extent predictive map of the roadkill risk, while separating exposure and danger and accounting for uneven survey effort. Because the initial aim of a multispecies comparative model quickly appeared intractable at the targeted spatial extent, we chose to focus on seven mammal species and functional groups that are regular victim of collisions with vehicles. These species are sufficiently common in our study area and large enough to expect that animals dead at the collision location will be detected in most instances and are usually recorded by road patrols. Furthermore, all species have a sufficiently large home range and broad habitat use within the survey area to act as surrogates for other species’ responses to roads. This has previously been shown for roe deer (Polak et al. 2014), which we focus on for demonstration in the main text.

To construct our spatial predictive map, we hypothesized that the roadkill risk was modulated by a hierarchical interaction between ecological predictors of roe deer densities and local, road-level factors expected to modulate collision danger. We therefore elaborated an integrated risk analysis structured by two hierarchically-organized spatial point process models, accounting for sampling error and informed by a parsimonious set of covariates chosen to represent species’ use of landscape and some key properties of roads. Although the literature and experts knowledge reports seasonal and interannual variations in the behaviour, use of space and casualties of some species (Raymond et al. 2021), public policies primarily require a mapping of the average roadkill risk over the long term. We therefore prioritized our effort on the spatial pattern, leaving temporal structure for subsequent studies. We eventually discuss the results considering policy expectations, ecological realism and current challenges for data sampling.

[Julien : peut être renforcer un peu le côté « nouveauté » du modèle afin de montrer qu’on en est à explorer de nouvelles solutions, pas de faire de la routine. Il faut que le lecteur soit incité à s’intéresser au pattern d’ensemble et à la méthode plutôt qu’aux espèces particulières qu’on étudie]

**Materials and methods**

*Study area*

We chose Brittany, France’s westernmost region, as our study area (Fig.1). This choice was primarily dictated by policy reasons, but the region also has some features that facilitated the construction of our model. First, being a peninsula, Brittany is well delineated in space, which helped matching the policy relevance of our modelling window with a biogeographically consistent unit. Second, topographical variation is sufficiently low to avoid crossing the altitudinal margins of the species encompassed in our study. Third, large roads in Brittany are well covered by an active network of road patrols and is well covered by amateur naturalists who report their records on a public database benefiting from an efficient reporting and curation system.

We restricted our analysis to the roads on which we had roadkill data, corresponding to a network of 1590 km with 1250 km of 2 x 2 lanes roads (max. speed 110 km.h-1) and 340 km of 1 x 1 lanes roads (max. speed 90 km.h-1). We did not cover local service roads. Our sample was thus biased towards larger, high-speed, high-traffic roads, which may not be those where roadkills are at their highest levels. However, these large roads are prioritized by public policies for the reduction of ecological impacts of infrastructures, making them the most meaningful for a first attempt to construct a predictive spatial risk model.

Une image contenant texte, carte, diagramme

Le contenu généré par l’IA peut être incorrect.

**Figure 1.** (a) Study area and its location in France, with the roads included in this study ; (b-d) three predictors (resolution 1 x 1 km, scaled units, darker color means higher values). All the maps for raw predictors and species records can be found in Appendix SM1 .

*Species selection*

We initially focused on ten mammal species that are reasonably well known both by naturalists and road patrollers, and with sufficiently large body sizes to expect detection of killed individuals during non-dedicated patrols. We discarded other species due to sparse records and strong spatial patterns unrelated to biology. The sample included wild boar *Sus scrofa*, red fox *Vulpes vulpes,* european badger *Meles meles,* roe deer *Capreolus capreolus*, red deer *Cervus elaphus*, European hare *Lepus europaeus*, pine marten *Martes martes,* house marten *Martes foina,* least weasel *Mustela nivalis* and stoat *Mustela erminea.* We grouped the latter four as mustelids (*Mustelidae*, referred to as one of the “species” in the text for simplicity) due to obvious confusions among species or random assignations by patrollers within this family. We performed this pooling only after data curation and exploration, and after having performed preliminary models with all species separated that suggested obvious flaws based on our knowledge of species’ ecologies. We thus eventually worked on seven species, among which two pose serious public safety threats on roads (wild boar and roe deer) and one has a special emblematic value for the study area (stoat, included among *Mustelidae*).

No regional population density estimates are published even for game species in this region (wild boar and roe deer), but all are widespread and common in the study area. Habitat-related variations in the detection of alive individuals exist, especially since all species occur on a wide range of habitats encompassing urban peripheries, bocages and woodlands. This may bias the estimation of spatial variation in presence probability unless opportunistic data can be paired with protocoled surveys with replication, which do not exist currently with a sufficient representativity in the study area. However, exploratory maps of raw data and model predictions showed patterns that matched regional experts’ assessment of the seven species’ distributions. Our model-based predictions are thus to be understood as an initial draw in the current, incomplete state of knowledge, a well-known limitation related to the Wallacean bias which can be overcome only by increasing sampling effort (Hortal et al. 2015).

*Exposure layer - alive animals distribution data*

The only database on mammal distributions available in Brittany is dominated by opportunistic records (https://atlas.gmb.bzh/), and existing protocoled counts were too sparse to be useful for spatial analysis even for the best monitored species like wild boar and roe deer. Hence, we chose to exploit these opportunistic data, assuming that their coverage and quantity for the seven species considered in our study balanced the lack of standardization, although this can hardly be demonstrated. Search bias is however probably limited by the fact that we only considered common, non-flagship species that are not specifically sought after by naturalists. As a preliminary verification, we subjected raw data maps and model-based predictions to regional experts, but we did not attempt for a more formal modelling of detectability patterns, which would have inflated model complexity and computation demand without any expectable improvement in outcome quality (see also the ‘*Sampling effort data*’ section below). This means that in spite of our confidence in experts’ validation of the main trends, the uncertainty estimates reported in our results are probably underestimated to an unquantifiable extent. An additional motivation for employing opportunistic and patrol data in spite of their limitations is the increasing interest of policymakers for the potential of these data to assist monitoring and decision, due to their low acquisition cost and high spatial and temporal coverage (Dobson et al. 2020).

We retrieved opportunistic geolocated records for the eight species from the ‘Faune Bretagne’ and ‘Faune Anjou’ databases within the ‘Faune France’ portal (<https://www.faune-france.org/>) developed under the Biolovision architecture (<https://data.biolovision.net/>). [LPO : indiquer s’il faut citer des refs spécifiques]. This database is fed by volunteer naturalists who gather geolocated species occurrence records opportunistically through a smartphone interface. The data are subsequently curated for location accuracy and identification. We selected only data on alive animals with the highest spatial precision and confidence level. We restricted the data to the period from 2015 to 2020, for which roadkill data were retrieved with a reasonably homogeneous spatial and temporal effort throughout the region (see the *Roadkill data* section below). We pooled records from all years and seasons to obtain a global picture of species’ regional distributions (see Introduction). Raw exposure data can be found in Appendix SM1 - 1.

*Exposure layer - quantification of sampling effort*

We quantified sampling heterogeneities only for the exposure layer (*i.e.*, alive animal records) since effort was considered sufficiently homogeneous across roadkill data to be ignored in an analysis for which the target is to predict relative spatial patterns rather than the absolute roadkill risk. Since we had no protocoled data to rely on for a proper calibration of sampling effort in our opportunistic spatial point pattern, we used a set of four proxies of sampling effort identified through a consultation with the database curators based on an initial proposal derived from a previous study using comparable data (De Solan et al. 2019). We extracted these proxies from the entire database of live animals from 2015 to 2020, all species confounded, on a 5 x 5 km square grid:

* total number of records (reflects sampling pressure on a grid cell);
* number of dates with records (reflects sampling frequency on a grid cell);
* number of species recorded (reflects the diversity and motivation of naturalists passing through a grid cell);
* number of observers (reflects heterogeneity in individual strategies of data collection).

These four variables were correlated (from r² = 0.47 between the number of species recorded and the number of observers to r² = 0.86 between the number of dates and the number of records). We thus summarized sampling effort as the first axis of a principal component analysis on these four variables (PC1, Fig.1d), which accounted for 74.6% of the total inertia (14.6% for the next axis). This principal component was negatively correlated to all sampling effort proxies (Appendix SM1 - 2). We defined as a synthetic, regularized proxy of sampling effort for each cell.

*Exposure layer – environmental predictors*

We computed all exposure predictors on a 1 x 1 km grid covering all the study area (Appendix SM1 - 3 for maps). We retrieved the surface cover of four land cover types (OSO 2020 : <https://www.theia-land.fr/ceslist/ces-occupation-des-sols/> , native resolution 10 x 10 m, retrieved march 2021) : agriculture (OSO codes 5 to 12), forest (OSO codes 16 and 17), wetland (OSO code 23) and artificialized areas (OSO codes 1 to 3, dominated by the urban network). These categories capture the main compositional features of landscapes in our study area and correlate with less common habitats, such as meadows or moorlands, which were too under-represented to be included. We also computed the total length of hedgerows per pixel from a public database maintained by the Insitut Géographique National (IGN -, BD-TOPO: <https://geoservices.ign.fr/telechargement>, retrieved in march 2021 – vector layer HAIE.shp). All these variables were inter-correlated. We therefore summarized them in the four first axes of a principal component analysis, which respectively accounted for 37%, 23%, 19% and 16% of the total explained variance (last axis 5%, excluded). The first axis (PC1) was a closure gradient separating agricultural landscapes from forests; PC2 ranged from urban to rural areas; PC3 was related to hydrology; PC4 discriminated bocages from intensive agriculture (see PC plots in Appendix SM1 - 4).

*Roadkill data*

Roadkill records have been gathered since 2014 by the patrols of the Direction Interdépartementale des Routes Ouest (DIR), the public administration in charge of safety and maintenance on the public road network covered by our study. The whole road network is entirely surveyed nearly every day, but the number of patrols per day varies to an unknown extent depending on service needs. Patrols are operated in a car circulating at a regular speed varying from 60 to 100 km.h-1 according to road type (lower speeds would cause safety issues). Non-specialized DIR officers identify all killed animal seen on the road and record their position from the nearest kilometric marker (no geolocation), which implies that the precision of records may vary over ±200m depending on the patrol speed and officers experience. Records of roadkills are mandatory during patrols but do not form their primary mission, limiting the risk of over-reporting, biased search patterns or other limitations that arise with dedicated patrols (Keane et al. 2011). Since these data are not curated, we screened them for obvious identification errors, misspelling or inadequate localisation. We deleted all unidentified ungulates, although most of them are probably roe deers, and all unidentified small mammals. Apart from mustelids (see above), none of the species chosen for our study was evidently concerned by such errors.

An unknown proportion of roadkills comes undetected (*e.g.* wounded animals dying far from the road, killed animals gathered by drivers after collision, carcasses hidden by vegetation), but we have no reason to believe that they are structured spatially in a way that could lead to confounding effects in analyses. The grain at which this assumption holds would however deserve further exploration with an experimental design generating protocoled data, because carcass detection and degradation rates can still be modified by traffic, vehicle speed and other spatially varying factors (Bénard et al. 2024). Accumulating such data is out of reach in a reasonable time and effort given the regional extent of the region, such that public policy requests for regional-scale models will probably have to live with this type of uncertainty until dedicated projects have been set up. For our own purpose, we accepted the data on the basis that they did not retain any spatial structure suggesting heterogeneities in detections of roadkills or species assignment.

**Table 1.** Descriptive summary of the predictors of the exposure, danger and sampling layers. The resolutions in italics are those of the data extraction

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **layer** | **variable** | **mean** | **sd** | **min** | **max** | **pixels>0 (%)** |
| exposure  *(1 x 1 km pixels)* | Urban cover, ha | 5,00 | 10,16 | 0,00 | 97,65 | 54,09 |
| Forest cover, ha | 9,92 | 16,69 | 0,00 | 100,00 | 54,47 |
| Agriculture cover, ha | 18,37 | 22,91 | 0,00 | 96,58 | 53,65 |
| Wetlands cover, ha | 0,01 | 0,04 | 0,00 | 0,69 | 11,27 |
| Hedgerows, m.ha-1 | 34,31 | 38,83 | 0 | 206,36 | 53,55 |
| danger | distance to water, m *(500 x 500 m pixels)* | 164,97 | 704,50 | 0,00 | 10771,67 | 8,52 |
| distance to vegetation, m *(50 x 50 m pixels)* | 3,17 | 23,60 | 0,00 | 1958,78 | 6,76 |
| speed, km.h-1 *(1 x 1 km pixels)* | 2,07 | 14,12 | 0,00 | 125,00 | 2,17 |
| traffic, cars.day-1 *1 x 1 km pixels)* | 509,67 | 4112,32 | 0,00 | 111128,41 | 2,16 |
| sampling | sampling pressure *(total nb. of records in database)* | 2775,82 | 5409,47 | 0,00 | 65316,00 | 99,28 |
| sampling frequency *(nb. of dates in database)* | 297,25 | 378,45 | 0,00 | 3704,00 | 99,28 |
| sampling diversity *(nb. of animal species in database)* | 131,65 | 83,90 | 0,00 | 1003,00 | 99,28 |
| observer diversity *(nb. of observers in database)* | 59,27 | 66,65 | 0,00 | 631,00 | 99,28 |

*Danger layer – environmental predictors*

The immediate environment around a road has an attracting or repulsing effect that may inflate crossing probabilities, concentrate animals on favourable roadsides or trigger collision-prone behaviours. We therefore included the proximity of roads to vegetation and water as predictors in the danger layer, considering that they form proxies of ecological corridors, resource availability and suitable local habitats (Benitez-Lopez et al. 2010). We measured these two variables from vector layers produced by the IGN (BD-TOPO, shapefiles HAIE.shp - ZONE\_DE\_VEGETATION.shp - TRONCON\_HYDROGRAPHIQUE.shp), that we rasterized at a 500 x 500 m resolution (vegetation) and 50 x 50 m resolution (water). The difference in resolution is motivated by the *ad hoc* expectation that vegetation has a higher range of influence than water for the species included in our study, all terrestrial mammals related to some extent to hedgerows and woods. To obtain the distance between a road and the nearest vegetation patch, we computed the nearest distance between each pixel centroid and the nearest vegetation polygon or hedgerow. We processed similarly for distance between roads and water, accounting for all major streams and rivers (order 4 or higher in the IGN BD-TOPO nomenclature, excluding temporary waterbeds and minor streams).

We extracted vehicle speed from a vector layer (IGN BD-TOPO, shapefile TRONCON\_DE\_ROUTE.shp). These data provide the average speed of light vehicles per road sections, modelled from the type of roads and their location in the regional road network. We averaged speed in a 1 x 1 km raster and converted it as a binary variable with a cutoff at 90 km.h-1, which represents the main threshold between local service roads and major axes in our study area. In many instances, this speed threshold also differentiates roads with no or little infrastructure likely to prevent crossings (*e.g.* fences, guardrails or central reservations), on which no direct data are available.

We retrieved data on road traffic through the DIR from the IRIS software (<https://dreal.applis-bretagne.fr/iris/calcul_trafic/connexion.php>, accessed march 2021) [CEREMA attention : le lien est HS (21/08/2024), insérer le bon lien, que je n’ai pas retrouvé], which interpolates vehicle counts through inductive loops from automated counting stations spread all over the road network [CEREMA : si vous pouvez ajouter quelques precisions : nombre de capteurs, espacement, références, etc…]. The most relevant data layer for our dataset was the interpolated traffic compiled during year 2020, expressed in vehicles per day including all types of vehicles and averaged in a 1 x 1 km raster grid [CEREMA : quelle est la résolution native?]. Although year 2020 was peculiar due to the COVID-19 pandemics, screening the raster with regional experts revealed that the spatial pattern conformed to expectations, suggesting that the relative variations in traffic were preserved although the absolute traffic was lower than usual.

*Risk analysis model*

We treated alive animal sightings and roadkill records as two collections of spatial points, in which each point represented either the presence of a living animal, , or an animal killed by a vehicle collision, , for a given species at a given location . These data can be modelled in a point process framework (Renner et al. 2015), which predicts the number of points per unit area through intensity functions for and for . A fundamental implementation of this model is the inhomogeneous Poisson process, which assumes that the intensity is spatially heterogeneous and that, conditional on covariates, points are independently distributed (Baddeley et al. 2015). In this study, we move beyond this independence assumption by employing a log-Gaussian Cox process (hereafter denoted as LGCP). This model allows for spatial interactions between points (*e.g.*, attraction or repulsion) unexplained by covariates alone, by incorporating a Gaussian random field (hereafter denoted as GRF) into the intensity function of the point process (Møller et al. 1998).

The intensity of alive individuals stems from the product of the true distribution of a given species, which defines the exposure of the risk model , and the probability of location to be sampled, such that:

[Julien fais un cross-check des équations STP . J’ai divisé la première en 2 pour faire des blocs d’équations moins gros, fluidifier la lecture et éviter les répétitions. OK ?]

(1)

We represented sampling effort through the covariate , described above and in Table 1 and modelled with habitat covariates and a spatial random effect:

(2)

where is the intercept, , …, are the coefficients associated with spatial covariates derived from the principal component analysis on land cover types, and is the coefficient related to the covariate quantifying sampling effort. The parameters and correspond to the range and standard deviation, respectively, of the Matérn covariance function used in the GRF.

Following the risk analysis framework, the intensity of roadkills stems from the exposure, , and the danger layer, :

(3)

We modelled as a function of road characteristics and a spatial random effect:

where is the intercept, and , …, are the coefficients associated with speed limits, traffic intensity, distance to watercourses, and distance to vegetation, respectively. In equation (4), we chose not to include exposure as an offset, as is frequent in risk models, to allow for a quantitative assessment of the impact of exposure on the spatial distribution of the risk (performed by ). The parameters and represent the range and standard deviation, respectively, of the Barrier covariance function used in the GRF. This Barrier model constrains spatial correlations to propagate along the road network (Bakka et al. 2019).

*Model implementation*

We implemented the model in INLA version 24.06.27 under R version 4.4.3 (Rue et al. 2009). INLA is a Bayesian inference method based on deterministic Laplace approximations, which contrasts with stochastic simulation-based approaches such as MCMC. This approach enables fast yet accurate estimation, particularly well-suited for spatial models through the SPDE framework (Lindgren et al. 2011). We conducted inference independently for each species. We achieved spatial discretization using a Voronoï mesh composed of cells with a mean area of 3.46 km² (median: 3.38 km²; 95% CI: [2.06 km², 5.33 km²]). The inference workflow was structured as follows:

1. *estimation of the spatial parameters:* estimate the range and standard deviation of the Matérn covariance function for the exposure model;
2. *fitting the E layer:* fit the full exposure model while fixing the parameters of the GRF to the previously estimated values;
3. *posterior sampling:* draw 100 posterior samples of the predicted exposure from the exposure model;
4. *risk model estimation:* for each posterior sample of , fit the corresponding risk model to estimate danger .

Une image contenant texte, capture d’écran, cercle, Police

Le contenu généré par l’IA peut être incorrect.

**Figure 2.** Diagram of the risk model. [Julien : revérifier cette figure stp]

*Model implementation*

We implemented the model in INLA version 24.06.27 under R version 4.4.3 (Rue et al. 2009). INLA is a Bayesian inference method based on deterministic Laplace approximations, which contrasts with stochastic simulation-based approaches such as MCMC. This approach enables fast yet accurate estimation, particularly well-suited for spatial models through the SPDE framework (Lindgren et al. 2011). We conducted inference independently for each species. We achieved spatial discretization using a Voronoï mesh composed of cells with a mean area of 3.46 km² (maps in SM1 – 1; median: 3.38 km²; 95% CI: [2.06 km², 5.33 km²]). The inference workflow was structured as follows:

1. *estimation of the spatial parameters:* estimate the range and standard deviation of the Matérn covariance function for the exposure model;
2. *fitting the E layer:* fit the full exposure model while fixing the parameters of the GRF to the previously estimated values;
3. *posterior sampling:* draw 100 posterior samples of the predicted intensity surface from the exposure model;
4. *risk model estimation:* for each posterior sample of , fit the corresponding risk model.

*Fitting and prediction assessment*

We assessed goodness of fit independently for the exposure model and for each risk model corresponding to a posterior sample. We considered the Area Under the Curve (AUC), the True Skill Statistic (TSS), and the correlation between observed and predicted counts. We conducted these assessments on spatial grids with resolutions of 20 km² for living animals and 10 km² for roadkills, respectively.

We assessed the robustness of predictions of the roadkill risk through a spatial cross-validation procedure. This involved removing a subset of the data, re-estimating model parameters, and predicting the held-out data. For each replicate, we draw a random point within the spatial domain and excluded all data within a 30 km radius for prediction. We performed a total of 100 replicates. We evaluated prediction quality using the AUC, the TSS, the Root Mean Square Error (RMSE), and the correlation between observed and predicted counts.

**Results**

*Factual figures on the data*

*Model performances*

Model fit was satisfactory overall (Table 2 – graphical overview in SM1 - 5), considering the relatively coarse resolution of spatial predictors in both model components, uncertainties inherent to opportunistic data, and the linear shapes imposed to covariates. Despite these constraints, AUC values exceeded 0.8 for all species. TSS values were expectedly lower but remained above 0.6 and 0.7 for most species in the exposure and risk model, respectively. Correlations between observed and fitted counts were higher and less variable for the risk layer (>0.8 for most species) than for the exposure layer (>0.7).

The cross-validation procedure implied the removal and prediction of 10% of the data per species on average. The model demonstrated good predictive performance (Table 3 – graphical overview in SM1 - 5), indicating the robustness of the inference procedure and suggesting limited overfitting for most species. Cross-validation outcomes varied more noticeably across replicates in Red deer, indicating a higher sensitivity to data availability than in other species which was coherent with the fit evaluation and the low number of roadkill records (Table 2),

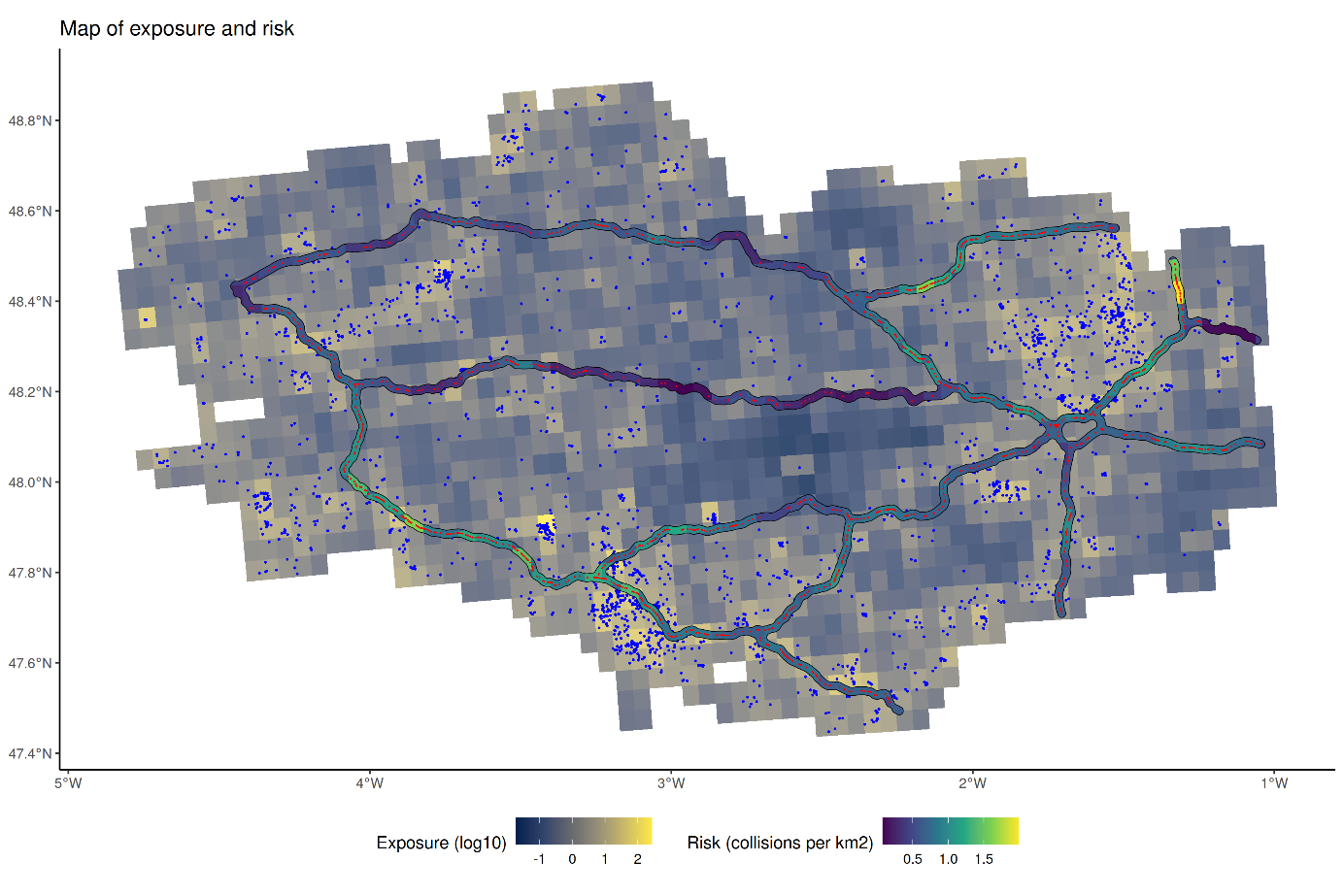
Fig. 3 illustrates the predictions of the exposure and risk models for Roe deer (other species in Appendix SM1 – XX). Spatial variations in the intensity of the exposure layer were coherent with raw data and are consistent with expert knowledge of the species’ distribution and abundance in the region. Nevertheless, the absence of any regional-level systematic atlas and estimates of population sizes prevents any formal validation. The predicted spatial patterns of roadkills did not depart strongly from the raw data in roe deer, suggesting that the spatial random effect was able to account for residual spatial variation unexplained by the covariates.

**Table 2** Sample sizes per species (N) and indicators of model fit. For the danger layer, the distribution of fit over 100 draws in the exposure layer is indicated with quantiles (median ; [IC 95%]).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Species** | **Exposure model** | | | | **Risk model** | | | |
| **N** | **AUC** | **TSS** | **COR** | **N** | **AUC** | **TSS** | **COR** |
| *Wild boar* | 212 | 0.90 | 0.66 | 0.70 | 730 | 0.93 [0.92,0.93] | 0.72 [0.71,0.74] | 0.93 [0.93,0.94] |
| *Red fox* | 1154 | 0.90 | 0.61 | 0.86 | 2360 | 0.98 [0.98,0.98] | 0.82 [0.81,0.82] | 0.91 [0.91,0.91] |
| *European badger* | 253 | 0.93 | 0.75 | 0.71 | 1158 | 0.96 [0.79,0.80] | 0.79 [0.79,0.80] | 0.87 [0.87,0.88] |
| *Roe deer* | 3518 | 0.90 | 0.63 | 0.86 | 1554 | 0.97 [0.97,0.97] | 0.81 [0.80,0.82] | 0.89 [0.89,0.89] |
| *European hare* | 1087 | 0.92 | 0.68 | 0.89 | 177 | 0.89 [0.88,0.89] | 0.66 [0.63,0.68] | 0.75 [0.74,0.76] |
| *Red deer* | 75 | 0.99 | 0.95 | 0.69 | 34 | 0.95 [0.93,0.97] | 0.76 [0.74,0.84] | 0.71 [0.68,0.74] |
| *Mustelids* | 390 | 0.84 | 0.51 | 0.57 | 1197 | 0.97 [0.96,0.97] | 0.81 [0.80,0.82] | 0.90 [0.90,0.91] |

**Table 3** Proportion of removed data and indicators of model prediction from the cross-validation procedure. The distribution of metrics over 100 replicates is indicated with quantiles (median ; [IC 95%]).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Species** | **% of removed data** | **AUC** | **TSS** | **RMSE** | **COR** |
| *wild boar* | 0.11 [0.02,0.17] | 0.89 [0.74,0.99] | 0.70 [0.50,0.97] | 1.32 [0.68,2.39] | 0.58 [0.27,0.92] |
| *red fox* | 0.10 [0.04,0.16] | 0.98 [0.83,1.00] | 0.90 [0.65,1.00] | 0.67 [0.50,1.31] | 0.77 [0.62,0.86] |
| *european badger* | 0.11 [0.03,0.17] | 0.95 [0.89,1.00] | 0.85 [0.69,1.00] | 0.78 [0.55,0.97] | 0.74 [0.56,0.84] |
| *roe deer* | 0.10 [0.05,0.16] | 0.96 [0.86,1.00] | 0.86 [0.62,1.00] | 0.71 [0.46,1.24] | 0.75 [0.57,0.89] |
| *European hare* | 0.10 [0.04,0.16] | 0.79 [0.63,1.00] | 0.56 [0.37,1.00] | 1.78 [1.13,2.62] | 0.44 [0.20,0.85] |
| *Red deer* | 0.11 [0.04,0.17] | 0.83 [0.59,0.98] | 0.73 [0.45,0.98] | 4.25 [1.68,7.77] | 0.26 [0.00,0.61] |
| *Mustelids* | 0.11 [0.03,0.16] | 0.97 [0.81,1.00] | 0.86 [0.62,1.00] | 0.88 [0.60,1.40] | 0.70 [0.47,0.86] |

**

**Figure 3.** Map of model predictions for the exposure and risk model along with observations of living animals (blue dots) and roadkill (red dots) for Roe deer. See appendix SM2 for similar plots on the other species.

*Effect of covariates*

In the exposure layer, most species were predominantly associated with wooded areas (Fig.4a). Wild boar, Roe deer, and European hare were mostly found in rural zones, while records of red fox were positively correlated with urban areas (Fig.4b). Unlike Red deer, Red fox and Roe deer showed positive associations with the density of watercourses (Fig.4c). Most species were found predominantly in bocage landscapes, except for Red deer, which was primarily observed in open field areas (Fig.4d). Finally, all species’ occurrence intensities were positively related with sampling effort, suggesting that the model was effective at capturing spatial variations in sampling effort (Fig.4e). [Julien, à clarifier : dans ce paragraphe les effets sont montrés à sampling constant (= exposure) ou c’est l’intensité des observations?]

Interestingly, exposure was uncorrelated with roadkill risk in all species (Fig. 5a), suggesting that roadkills are not adequate surrogate of species’ distributions and that local factors predominantly shape the danger layer. Speed increased risk for all but Red deer and Wild boar (Fig. 5b). Wild boar collisions occurred in areas with more intense traffic, whereas roadkill risk for European badger and hare were negatively correlated with traffic intensity (Fig. 5c). Distance to water and proximity to vegetation had idiosyncratic effects with no clear biological interpretation (Fig. 5d and 5e).

Une image contenant texte, diagramme, capture d’écran, nombre

Le contenu généré par l’IA peut être incorrect.

**Figure 4.** Effect sizes and 95% credibility intervals for the exposure model. All variables were scaled to have a mean of 0 and a standard deviation of 1. Blue, black, and red bars represent positive, non-significant, and negative relationships, respectively, based on whether the posterior probability that the corresponding parameter was greater (or lower) than 0 exceeded 0.95.

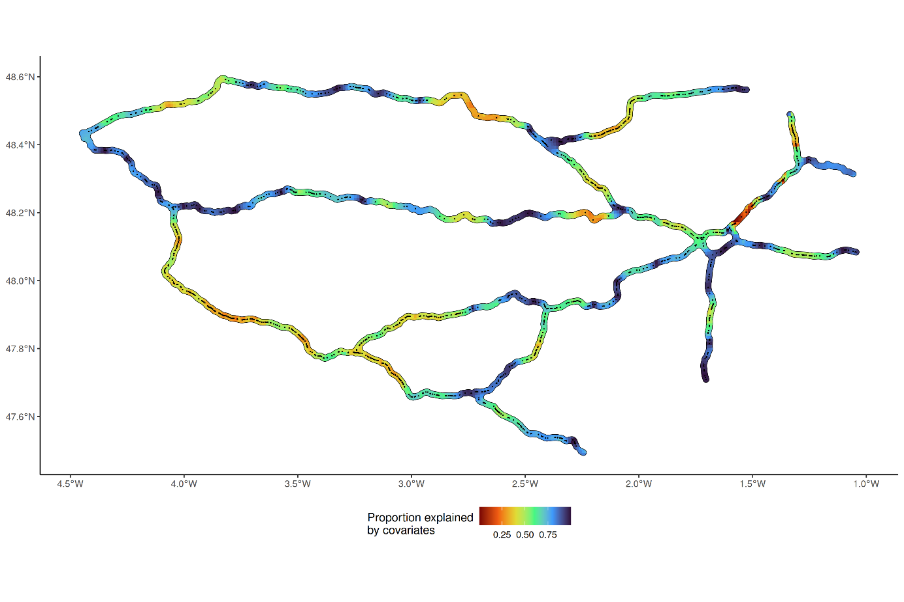
Une image contenant texte, capture d’écran, diagramme, Plan

Le contenu généré par l’IA peut être incorrect.

**Figure 5.** Effect sizes and 95% credibility intervals for the risk model based on the 100 posterior medians – one for each posterior sample of the exposure model. All variables were scaled to a 0 mean and unit standard deviation. Blue, black, and red bars represent positive, non-significant, and negative relationships, respectively, based on the mean of posterior probabilities over the 100 estimates.

*Proportion of roadkill variability explained by covariates*

An interesting feature of the model lies in its ability to map the proportion of variability in roadkill risk explained by the covariates. Road segments with high residual variability reflect the failure of covariates to account for the observed number of roadkills (in red on Fig. 6a). The proportion of explained variability decreased with the number of observed roadkills (Fig. 6b), indicating that our model was good at predicting accurately the main regional risk gradient, but less so hotspots of increased mortality. Segments with low explained variability thus point road sections where dedicated sampling protocols would be required to determine local factors that increase the collision risk.

**

Une image contenant capture d’écran, ligne, texte, diagramme

Le contenu généré par l’IA peut être incorrect.

**Figure 6.** Proportion of variability in roadkill intensity explained by the five covariates (Exposure, speed, traffic, distance to water and distance to vegetation). (a) red regions indicate that the covariates poorly explained the pattern of roadkills. Records are indicated by the black dots. (b) Relationship with the observed number of roadkill, with a parametric linear regression [Julien c’est bien une regression linéaire, pas un smoother qui a l’air linéaire?] and its 95% confidence interval. These figures refer to Roe deer, see appendix SM1 - XX for similar plots on the other species. [+ Julien est ce que ce ne serait pas plus pertinent de renverser la figure ?]

**Discussion**

Opportunistic records have become a key resource for biodiversity assessment [REF], but the biases inherent to the absence of sampling protocol require cautious modelling to avoid misleading predictions in spatial or temporal models [REF]. This is specifically critical when operational decision on ecological management relies on model-based spatial planning [REF]. Here, we focused on the wildlife roadkill risk both due to its multiple consequences for both biodiversity conservation and public safety, and because assessment of its spatial distribution at a regional scale can hardly be performed on the basis of protocoled data only. We showed that our risk model structure applied to a regional road network returned biologically realistic patterns, with satisfactory fit and spatial predictive performance. As compared with most pre-existing statistical predictions of regional roadkill patterns, our model structure clarifies the respective roles of covariates associated with exposure and danger, and accounts for sampling heterogeneities and residual spatial variation in a more realistic way.

A key outcome was the low correlation between the wildlife roadkill risk and species’ occurrence intensity patterns predicted in the exposure layer. This revealed that wildlife-vehicle collisions arise through more complex processes than a random sampling of species’ spatial distributions, a result further enforced by the clear role of speed and other covariates on the danger layer. This conclusion tempers previous studies that proposed roadkills as a pragmatic sampling method to assess the distribution of elusive or rare species [REF]. Although a roadkill sighting proves the presence of a given species on a given location, making it potentially effective to assess extent-of-occurrence limits, spatial variations in presence probability may thus not be adequately reflected by the distribution of killed animals along roads. Reciprocally, predictive roadkill risk maps issued from models that confound exposure and danger into a single layer, as is commonly done [REF], may lead to incorrect inference on covariates and result in inadequate management recommendations. Although an explicit risk model design comes at the cost of increased statistical complexity and requires distribution data that may not always be available, we thus propose it as an efficient and reliable solution to construct spatial predictions of the roadkill risk for well-documented species, such as those used in our study.

Overall, our model performed well both in fit and prediction, especially when considering the coarse nature of the data used and the numerous sources of uncertainty implied by opportunistic and semi-protocoled sampling. This result enforces the credibility of our model to inform regional-level predictive mapping in spatial planning policies [REF], although this comes with several limitations that need to be understood by practitioners. First, no independent dataset was available to subject our model to the most adequate test of predictive performance possible [REF], such that our current evaluation of the model is conditional to the processes that govern data generation at the regional level. In particular, roadkill data were generated by non-specialist patrols that may not detect and record killed animals with the same efficiency as dedicated operators or expert naturalists. Second, we did not control for the factors that affect carcass detectability, which are spatially structured themselves and may thus bias our perception of spatial risk patterns [REF : Bénard]. This issue is probably the most serious one, especially for predictions relative to species that may not die immediately after collisions or that are quickly removed by drivers for safety reasons (such as deers or wild boar [REF]). The resulting bias can hardly be controlled without referring to the results of controlled experiments or repeated sampling protocols [REF], which currently lack taxonomic and spatial coverage to be useful in a regional-scale model. Third, factors that control collision danger and exposure are temporally structured by seasonality [REF] and inter-annual variations [REF], making the addition of a temporal component an obvious next step in the development of a fully operational risk model. At this stage, our exploration of methodological options revealed that a spatial-temporal risk model would require a better temporal coverage than available in our data and increased computation costs reaching the current limits of most users’ capacities. [Julien, est-ce que tu pourrais abonder ce paragraphe avec un peu (2-3 phrases de discussion et biblio plutôt orientée stats pour confronter ça à d’autres modèles publiés ou renforcer les arguments ?].

These limitations being granted, the predictions of intensity variations in the exposure layer and the risk patterns appeared reasonable given the current state of regional experts’ knowledge on the selected species. Covariate effects in the exposure layer conformed to known habitat preferences, except for some hardly detectable species such as badger, for which sampling effects may not be fully accounted for. Covariates associated to danger showed heterogeneous effects that mostly conformed to the literature and biological expectations. In particular, vehicle speed had a consistently positive effect on danger, as found in other contexts [REF]. Because all high-speed sections in our road network were fenced, this pattern suggests that fencing had a limited efficiency [voir ce qu’en dit la biblio]. The effects of traffic were more equivocal, perhaps because high traffic densities exert a repulsive effect on animals close to roads [REF]. The role of distance to vegetation was less easy to interpret because animal propensity to cross roads and carcass detection by operators both decrease close to vegetation [REF]. These ambiguities in covariates interpretations may turn problematic to elaborate risk mitigation solutions [REF]. In particular, previous studies insisted on the need to account for fine-grained spatial variation in habitat (such as vegetation height along roadsides [REF]), fencing [REF] and other pre-existing infrastructure. Unfortunately, these data currently do not exist with a sufficient level of updating in our study region, a deficiency that should be taken seriously since it conditions the possibility to increase the spatial resolution of roadkill risk predictions, which remain quite coarse in our model (although still better than in other regional studies [REF]). That said, the impact of possible inaccuracies in habitat effects on the predictive performance of our model seems limited, suggesting that the relatively rough proxies we used are sufficient if the aim is to produce reasonably accurate maps at a regional extent.

The mapping of spatial residuals revealed that predictive performance dropped in some restricted road segments, which were associated with high roadkill record densities. Our model was thus good at predicting the main spatial risk pattern over the region but failed at identifying mortality hotspots. This may either be because data resolution in the exposure layer was insufficient to reveal locally high presence intensities, or because we missed covariates explaining hotspot occurrence [check refs sur hotspots]. However, spatial non-stationarity in the processes that trigger hotspots may limit parsimonious improvements of the model and restrain the relevance of gathering a single, regionally consistent set of fine-grained covariates. Integrated models may ultimately allow joint predictions of regional and local risk patterns by unifying the potentials of coarse regional records and local protocoled surveys [REF], as long as spatial variation is adequately represented at both scales. Since this development is still far-off given current data availability, we suggest that spatial residuals from models built on opportunistic records may first be used to prioritize protocoled sampling on specific localities that deserve increased understanding of collision factors.

*Operational implications (public policies)*

Our conclusions are positive on the usability of opportunistic data to provide biologically relevant and relatively robust predictive maps of the wildlife roadkill risk for regional-level spatial planning. However, pretending that this model is fully adequate for operational use by road managers confronted to roadkill collisions would be an overstatement. Since models can only be as good as the underlying data [REF], we propose that a first iteration of a risk model solely based on non-protocoled records should be used for sampling prioritization rather than decision making. This step should lead to the planification of protocoled surveys at locations optimized for the spatial coverage of areas in which opportunistic data reveal insufficient, and to identify ambiguities in pattern interpretations that suggest detection biases or deficiencies in covariates. This step is especially relevant in contexts similar to our roadkill risk results, in which mortality hotspots seem to respond to finer-grained determinants than the average regional pattern. Hence, we suggest that no model complexification would really be useful if not associated with improvement in data collection in the prospect of building integrated risk models. Furthermore, quantitative predictive assessment and expert-based advice are complementary and should be used in association for model validation. Beyond local experts’ knowledge of species distributions and their useful intuition on some spatial patterns, involving them in the model building process increased the credibility of our predictions and acceptation of the underlying technical complexity. As a whole, we suggest that biologically realistic modelling frameworks, such as the risk model presented here, are both a way to improve transparency on results, identify possibilities and issues associated with opportunistic records and inform prioritization of protocoled data sampling for more operational biodiversity surveys.

**Supporting Information**

**Appendix SM1 :** Supplementary figures.

**References**

Ahmed, S. S., Cohen, J. and Anastasopoulos, P. C. 2021. A correlated random parameters with heterogeneity in means approach of deer-vehicle collisions and resulting injury-severities. - Anal. Methods Accid. Res. 30: 100160.

Altwegg, R. and Nichols, J. D. 2019. Occupancy models for citizen-science data. - Methods in Ecology and Evolution 10: 8–21.

Ascensão, F., Yogui, D. R., Alves, M. H., Alves, A. C., Abra, F. and Desbiez, A. L. J. 2021. Preventing wildlife roadkill can offset mitigation investments in short-medium term. - Biological Conservation 253: 108902.

Baddeley, A., Rubak, E. and Turner, R. 2015. Spatial Point Patterns: Methodology and Applications with R. - Chapman and Hall/CRC.

Bakka, H., Vanhatalo, J., Illian, J. B., Simpson, D. and Rue, H. 2019. Non-stationary Gaussian models with physical barriers. - Spatial Statistics 29: 268–288.

Balčiauskas, L., Kučas, A. and Balčiauskienė, L. 2025. A Review of Wildlife–Vehicle Collisions: A Multidisciplinary Path to Sustainable Transportation and Wildlife Protection. - Sustainability 17: 4644.

Bénard, A., Bonenfant, C. and Lengagne, T. 2024. Traffic and weather influence on small wildlife carcass persistence time on roads. - Transportation Research Part D: Transport and Environment 126: 104012.

Benitez-Lopez, A., Alkemade, R. and Verweij, P. A. 2010. The impacts of roads and other infrastructure on mammal and bird populations: a meta analysis. - Biological Conservation 143: 1307–1316.

Benítez-López, A., Alkemade, R. and Verweij, P. A. 2010. The impacts of roads and other infrastructure on mammal and bird populations: A meta-analysis. - Biological Conservation 143: 1307–1316.

Bissonette, J. A. and Kassar, C. A. 2008. Locations of deer–vehicle collisions are unrelated to traffic volume or posted speed limit. - Human-Wildlife Conflicts 2: 122–130.

Boyle, S. P., Litzgus, J. D. and Lesbarrères, D. 2017. Comparison of road surveys and circuit theory to predict hotspot locations for implementing road-effect mitigation. - Biodivers Conserv 26: 3445–3463.

Coulson, G. and Bender, H. 2019. Roadkill mitigation is paved with good intentions: a critique of Fox et al. (2019). - Aust. Mammalogy 42: 122–130.

De Solan, T., Renner, I., Cheylan, M., Geniez, P. and Barnagaud, J.-Y. 2019. Opportunistic records reveal Mediterranean reptiles’ scale-dependent responses to anthropogenic land use. - Ecography 42: 608–620.

Denneboom, D., Bar-Massada, A. and Shwartz, A. 2024. Wildlife mortality risk posed by high and low traffic roads. - Conservation Biology 38: e14159.

Dobson, A. D. M., Milner-Gulland, E. J., Aebischer, N. J., Beale, C. M., Brozovic, R., Coals, P., Critchlow, R., Dancer, A., Greve, M., Hinsley, A., Ibbett, H., Johnston, A., Kuiper, T., Le Comber, S., Mahood, S. P., Moore, J. F., Nilsen, E. B., Pocock, M. J. O., Quinn, A., Travers, H., Wilfred, P., Wright, J. and Keane, A. 2020. Making Messy Data Work for Conservation. - One Earth 2: 455–465.

Fabrizio, M., Di Febbraro, M., D’Amico, M., Frate, L., Roscioni, F. and Loy, A. 2019. Habitat suitability vs landscape connectivity determining roadkill risk at a regional scale: a case study on European badger (Meles meles). - Eur J Wildl Res 65: 7.

Forman, R. T. T. and Alexander, L. E. 1998. Roads and Their Major Ecological Effects. - Annual Review of Ecology and Systematics 29: 207-C2.

Frangini, L., Sterrer, U., Franchini, M., Pesaro, S., Ruedisser, J. and Filacorda, S. 2022. Stay home, stay safe? High habitat suitability and environmental connectivity increases road mortality in a colonizing mesocarnivore. - Landsc. Ecol. 37: 2343–2361.

Geurts, E. M., Reynolds, J. D. and Starzomski, B. M. 2023. Turning observations into biodiversity data: Broadscale spatial biases in community science. - Ecosphere 14: e4582.

Gkritza, K., Baird, M. and Hans, Z. N. 2010. Deer-vehicle collisions, deer density, and land use in Iowa’s urban deer herd management zones. - Accident Analysis & Prevention 42: 1916–1925.

González-Suárez, M., Zanchetta Ferreira, F. and Grilo, C. 2018. Spatial and species-level predictions of road mortality risk using trait data. - Global Ecology and Biogeography 27: 1093–1105.

Grilo, C., Ferreira, F. Z. and Revilla, E. 2015. No evidence of a threshold in traffic volume affecting road-kill mortality at a large spatio-temporal scale. - Environmental Impact Assessment Review 55: 54–58.

Grilo, C., Koroleva, E., Andrášik, R., Bíl, M. and González‐Suárez, M. 2020. Roadkill risk and population vulnerability in European birds and mammals. - Frontiers in Ecology and the Environment 18: 323–328.

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. and González-Suárez, M. 2021. Conservation threats from roadkill in the global road network. - Global Ecology and Biogeography 30: 2200–2210.

Grilo, C., Neves, T., Bates, J., le Roux, A., Medrano-Vizcaíno, P., Quaranta, M., Silva, I., Soanes, K. and Wang, Y. 2025. Global Roadkill Data: a dataset on terrestrial vertebrate mortality caused by collision with vehicles. - Sci Data 12: 505.

Guisan, A. 2017. Habitat Suitability and Distribution Models. - Cambridge University Press.

Gunson, K. E., Mountrakis, G. and Quackenbush, L. J. 2011. Spatial wildlife-vehicle collision models: A review of current work and its application to transportation mitigation projects. - Journal of Environmental Management 92: 1074–1082.

Gurumurthy, K. M., Bansal, P., Kockelman, K. M. and Li, Z. 2022. Modelling animal-vehicle collision counts across large networks using a Bayesian hierarchical model with time-varying parameters. - Anal. Methods Accid. Res. 36: 100231.

Ha, H. and Shilling, F. 2018. Modelling potential wildlife-vehicle collisions (WVC) locations using environmental factors and human population density: A case-study from 3 state highways in Central California. - Ecol. Inform. 43: 212–221.

Haddad, N. M. 2015. Corridors for people, corridors for nature. - Science 350: 1166–1167.

Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M. and Ladle, R. J. 2015. Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. - Annual Review of Ecology, Evolution, and Systematics 46: 523–549.

Isaac, N. J. B., van Strien, A. J., August, T. A., de Zeeuw, M. P. and Roy, D. B. 2014. Statistics for citizen science: extracting signals of change from noisy ecological data. - Methods Ecol Evol 5: 1052–1060.

Jackson, N. D. and Fahrig, L. 2011. Relative effects of road mortality and decreased connectivity on population genetic diversity. - Biological Conservation 144: 3143–3148.

Kang, W., Minor, E. S., Woo, D., Lee, D. and Park, C.-R. 2016. Forest mammal roadkills as related to habitat connectivity in protected areas. - Biodivers Conserv 25: 2673–2686.

Keane, A., Jones, J. P. G. and Milner-Gulland, E. J. 2011. Encounter data in resource management and ecology: pitfalls and possibilities. - Journal of Applied Ecology 48: 1164–1173.

Kušta, T., Keken, Z., Ježek, M., Holá, M. and Šmíd, P. 2017. The effect of traffic intensity and animal activity on probability of ungulate-vehicle collisions in the Czech Republic. - Safety Science 91: 105–113.

Laube, P., Ratnaweera, N., Wróbel, A., Kaelin, I., Stephani, A., Reifler-Baechtiger, M., Graf, R. F. and Suter, S. 2023. Analysing and predicting wildlife–vehicle collision hotspots for the Swiss road network. - Landsc Ecol 38: 1765–1783.

Laurance, W. F. and Balmford, A. 2013. A global map for road building. - Nature 495: 308–309.

Lima Santos, R. A., Santos, S. M., Santos-Reis, M., de Figueiredo, A. P., Bager, A., Aguiar, L. M. S. and Ascensao, F. 2016. Carcass Persistence and Detectability: Reducing the Uncertainty Surrounding Wildlife-Vehicle Collision Surveys. - PLoS One 11: e0165608.

Lindgren, F., Rue, H. and Lindström, J. 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. - Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73: 423–498.

Mair, L. and Ruete, A. 2016. Explaining Spatial Variation in the Recording Effort of Citizen Science Data across Multiple Taxa. - PLOS ONE 11: e0147796.

Malo, J. E., Suárez, F. and Díez, A. 2004. Can we mitigate animal–vehicle accidents using predictive models? - Journal of Applied Ecology 41: 701–710.

Martins, T., Rodrigues Freitas, S., Lupinetti-Cunha, A., Semensatto, D. and Hardt, E. 2024. Combining roadkill hotspots and landscape features to guide mitigation measures on highways. - Journal for Nature Conservation 82: 126738.

Meijer, J. R., Huijbregts, M. A. J., Schotten, K. C. G. J. and Schipper, A. M. 2018. Global patterns of current and future road infrastructure. - Environ. Res. Lett. 13: 064006.

Møller, J., Syversveen, A. R. and Waagepetersen, R. P. 1998. Log Gaussian Cox Processes. - Scandinavian Journal of Statistics 25: 451–482.

Morelli, F., Benedetti, Y. and Delgado, J. D. 2020. A forecasting map of avian roadkill-risk in Europe: A tool to identify potential hotspots. - Biological Conservation 249: 108729.

Pinto, T., Santos, S. M., Mira, A. and Sillero, N. 2023. Importance of water availability for amphibian roadkill in a mediterranean landscape. - Biodivers. Conserv. 32: 2513–2537.

Plante, J., Jaeger, J. A. G. and Desrochers, A. 2019. How do landscape context and fences influence roadkill locations of small and medium-sized mammals? - Journal of Environmental Management 235: 511–520.

Polak, T., Rhodes, J. R., Jones, D. and 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. and Perkins, S. E. 2021. Temporal patterns of wildlife roadkill in the UK. - PLOS ONE 16: e0258083.

Renner, I. W. and Warton, D. I. 2013. Equivalence of MAXENT and Poisson Point Process Models for Species Distribution Modeling in Ecology. - BIOM 69: 274–281.

Renner, I. W., Elith, J., Baddeley, A., Fithian, W., Hastie, T., Phillips, S. J., Popovic, G. and Warton, D. I. 2015. Point process models for presence-only analysis. - Methods Ecol Evol 6: 366–379.

Román, J., Rodríguez, C., García-Rodríguez, A., Diez-Virto, I., Gutiérrez-Expósito, C., Jubete, F., Paniw, M., Clavero, M., Revilla, E. and D’Amico, M. 2024. Beyond crippling bias: Carcass-location bias in roadkill studies. - Conservation Science and Practice 6: e13103.

Rue, H., Martino, S. and 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.

Russo, L. F., Barrientos, R., Fabrizio, M., Febbraro, M. D. and Loy, A. 2020. Prioritizing road-kill mitigation areas: A spatially explicit national-scale model for an elusive carnivore. - Diversity and Distributions 26: 1093–1103.

Rytwinski, T. and Fahrig, L. 2015. The Impacts of Roads and Traffic on Terrestrial Animal Populations. - In: Handbook of Road Ecology. John Wiley & Sons, Ltd, pp. 237–246.

Rytwinski, T., Soanes, K., Jaeger, J. A. G., Fahrig, L., Findlay, C. S., Houlahan, J., Ree, R. van der and Grift, E. A. van der 2016. How Effective Is Road Mitigation at Reducing Road-Kill? A Meta-Analysis. - PLOS ONE 11: e0166941.

Santos, S. M., Lourenço, R., Mira, A. and 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.

Santos, R. A. L., Mota-Ferreira, M., Aguiar, L. M. S. and Ascensão, F. 2018. Predicting wildlife road-crossing probability from roadkill data using occupancy-detection models. - Science of The Total Environment 642: 629–637.

Shilling, F., Perkins, S. E. and Collinson, W. 2015. Wildlife/Roadkill Observation and Reporting Systems (R VanderRee, DJ Smith, and C Grilo, Eds.). - John Wiley & Sons Ltd.

Shin, Y., Kim, K., Groffen, J., Woo, D., Song, E. and Borzée, A. 2022. Citizen science and roadkill trends in the Korean herpetofauna: The importance of spatially biased and unstandardized data. - Front. Ecol. Evol. in press.

Swinnen, K. R. R., Jacobs, A., Claus, K., Ruyts, S., Vercayie, D., Lambrechts, J. and Herremans, M. 2022. “Animals under wheels”: Wildlife roadkill data collection by citizen scientists as a part of their nature recording activities. - Nat. Conserv.-Bulgaria: 121–153.

Tiago, P., Ceia-Hasse, A., Marques, T. A., Capinha, C. and Pereira, H. M. 2017. Spatial distribution of citizen science casuistic observations for different taxonomic groups. - Sci Rep 7: 12832.

Valerio, F., Basile, M. and Balestrieri, R. 2021. The identification of wildlife-vehicle collision hotspots: Citizen science reveals spatial and temporal patterns. - Ecological Processes 10: 6.

van der Grift, E. A., van der Ree, R., Fahrig, L., Findlay, S., Houlahan, J., Jaeger, J. A. G., Klar, N., Madriñan, L. F. and Olson, L. 2013. Evaluating the effectiveness of road mitigation measures. - Biodivers Conserv 22: 425–448.

van der Ree, R., Jaeger, J. A. G., van der Grift, E. and Clevenger, A. 2011. Effects of Roads and Traffic on Wildlife Populations and Landscape Function: Road Ecology is Moving toward Larger Scales. - Ecology and Society in press.

Van der Ree, R., Smith, D. J. and Grilo, C. 2015. The ecological effects of linear infrastructure and traffic. - In: Handbook of Road Ecology. John Wiley&Sons. in press.

Wright, P. G. R., Coomber, F. G., Bellamy, C. C., Perkins, S. E. and Mathews, F. 2020. Predicting hedgehog mortality risks on British roads using habitat suitability modelling. - PeerJ 7: e8154.