ECOGRAPHY

Research

Can we model distribution of population abundance from wildlife–vehicles collision data?

Javier Fernández-López, José A. Blanco-Aguiar, Joaquín Vicente and Pelayo Acevedo

J. Fernández-López (https://orcid.org/0000-0003-4352-0252) ✉ (J.FernandezLopez@umb.edu), J. A. Blanco-Aguiar, J. Vicente and P. Acevedo, Inst. de Investigación en Recursos Cinegéticos (IREC), CSIC-UCLM-JCCM, Ciudad Real, Spain. JF-L also at: Dept of Biology, Univ. of Massachusetts Boston, Boston, MA, USA.

Ecography
2022: e06113
doi: 10.1111/ecog.06113

Subject Editor: Damaris Zurell
Editor-in-Chief: Miguel Araújo
Accepted 14 March 2022

Reliable estimates of the distribution of species abundance are a key element in wildlife studies, but such information is usually difficult to obtain for large spatial or long temporal scales. Wildlife–vehicle collision (WVC) data is systematically registered in many countries and could be used as a proxy of population abundance if the number of WVC in each territory increase with the population abundance. However, factors such as road density or human population should be controlled to obtain accurate abundance estimations from WVC data. Here, we propose a hierarchical modeling approach using the Royle–Nichols model for detection–non-detection data to obtain population abundance indices from WVC. Relative abundance and individual detectability were modeled for two species, wild boar *Sus scrofa* and roe deer *Capreolus capreolus* at 10 × 10 km cells in mainland Spain from WVC data using environmental, anthropological and temporal covariates. For each cell, a detection was annotated if at least one WVC was recorded at each month (used as survey occasion). The predicted abundance indices were compared with raw hunting statistics at region level to assess the performance of the modeling approach. Site specific covariates such as road density or administrative region and the month of the year, affected individual detectability, with higher WVC probability between October and December for wild boar and between April and July for roe deer. Wild boar and roe deer abundance can be explained by both, bioclimatic and land cover covariates. Abundance indices obtained from WVC data were significantly positively correlated with regional raw hunting yields for both species. We presented empirical evidence supporting that accurate wildlife abundance indices at fine spatial resolution can be generated from WVC data when individual detectability is considered in the modeling process.

Keywords: *Capreolus capreolus*, detection–non-detection data, road ecology, roe deer, *Sus scrofa*, ungulates, wild boar, wildlife abundance estimation

NORDIC SOCIETY OIKOS

www.ecography.org

Introduction

Reliable estimates of species distribution and abundance are key elements in wildlife studies (Jones 2011). They allow for well-informed conservation plans, but also wildlife management, pest control or assessing epidemiological risks (Vicente et al. 2019). However, this basic information is not always easy to obtain due either to logistical constraints or intrinsic difficulties inherent to the studied species (i.e. nocturnal, elusive species, inaccessible habitats, complex social structure, etc.; Pearce and Boyce 2006).

Wildlife vehicle collisions (WVC) are an important problem for both, motorist security and wildlife conservation (Gunson et al. 2011, Ascensão et al. 2021). Roadkills can be one of the most important conservation threats for many animal populations specially reptiles and amphibians (Rytwinski and Fahrig 2012), but also birds or endangered mammals (Garrote et al. 2018, Grilo et al. 2020). In addition, when fatalities involve large species such as wild ungulates they constitute a serious risk for drivers that could led in injuries or human losses (Bissonette et al. 2008). In most countries with a well-developed road system, these events are recorded on a systematic basis. Despite of these obvious negative effects, WVC also represent a useful source of information for study species biology. Although the major focus of road ecology has been to study the fact of roadkills (Barrientos et al. 2021), in the simplest case a WVC indicates the presence of a species in a specific time and location (Colino-Rabanal and Peris 2016). They also could be used as a proxy of population abundance if the number of WVC increases with the population abundance (Baker et al. 2004). This spatial and temporal relationship has been already proved in different groups such as reptiles and amphibians (D'Amico et al. 2015), marsupials (*Perameles gunnii*, Mallick et al. 1998), carnivores (*Vulpes vulpes*, Baker et al. 2004; *Mustela putorius*, Barrientos and Miranda 2012) or ungulates (*Odocoileus virginianus*, McCaffery 1973; *Cervus elaphus*, *Capreolus capreolus* and *Sus scrofa*, Saint-Andrieux et al. 2020), showing the broad range of species to which this index could be applied (but see Ascensão et al. 2019a). WVC data have some interesting features to study population abundance: for large species, much information is systematically recorded by administrations or insurance companies across many countries, producing long historical time series at broad spatial scales (Vanlaar et al. 2012). However, despite their potential, WVC have not been widely used in wildlife population monitoring mainly due to specific singularities that prevent them from being directly used as abundance indices. Covariates such as road type and density, traffic intensity or vehicle speed have been often proven as the most important factors affecting animal road fatalities (Gunson et al. 2011). Thus, all these singularities should be accounted for in order to use WVC data as a population abundance index.

In the last decades, a novel modeling framework that integrates species detectability to study species distribution and abundance has become popular in wildlife monitoring programs (MacKenzie et al. 2002). Hierarchical modeling approaches account for false negatives in data samples using repetitive surveys to model both, the observational process (species detectability) and the underlying ecological process of interest (species occurrence or abundance; Royle and Nichols 2003). This framework would allow to effectively control for those non abundance-related (or collision-specific) factors affecting WVC to obtain accurate population abundance estimates. Such approach offers a new source of information for wildlife monitoring that could be easily compared across space and time with other different and complementary sources (Santos et al. 2018).

In this study we evaluated the utility of WVC to obtain population abundance indices of two widely distributed ungulates in mainland Spain, wild boar *Sus scrofa* and roe deer *Capreolus capreolus*, which are the two wildlife species most involved in vehicle collisions in this country (Sáenz-de-Santa-María and Tellería 2015). The Royle–Nichols model (RN) was used to infer abundances from repetitive WVC data accounting for individual detectability due to road related covariates such as road type or density (Royle and Nichols 2003). Model predictions were then projected to national scale to explore their relationship with an independent broadly used abundance index for game species at large spatial scales, namely the hunting yields (HY, ENETWILD Consortium et al. 2019).

Material and methods

Study area

Our study area spanned mainland Spain (493 518 km^2), in the Iberian Peninsula, southwestern Europe. Continental Spain is divided in 15 autonomous communities (thereafter regions) which are subdivided in 47 provinces, corresponding to the level 3 of the Nomenclature of Territorial Units for Statistics from the European Union (NUTS3). Spain is crossed by 666 677 km of paved roads heterogeneously distributed (Ministry of Transport, Mobility and Urban Agenda 2021; Supporting information). Human population in mainland Spain is distributed unevenly across the country, with higher densities in coastal areas an around the capital city in central Spain, Madrid, with a mean of 94 inhabitants km^{-2} (Spanish Statistical Office – INE 2021).

The study area was divided in 10 × 10 km cell grid which was used to assess the relationship between WVC and covariate predictors (environmental variables, road density, etc.). A total of 5184 cells were used in our analyses. A finer cell resolution could be used to obtain more precise abundance estimates but led to problems linked to sample size (number WVC per cell) and thus it was discarded.

Wildlife vehicle collision data

Two species were selected for abundance modeling from WVC data: wild boar and roe deer. These are the wildlife species most reported in vehicle collisions in mainland Spain, with

35% of wildlife collisions involving wild boar and 24.9% in the case of the roe deer during 2017. Note that those species are the most reported because their impact for human road safety, given their size, but other species are probably more affected by road kills (D'Amico et al. 2015). Populations of these two ungulates have increased during the last decades mainly due to rural abandonment, and their distribution ranges have also expanded significantly (Acevedo et al. 2011, Massei et al. 2015).

Wildlife vehicle collision data from 2017 January to December were obtained from the Dirección General de Tráfico (DGT) of the Spanish Ministry of the Interior that compiled traffic reports produced by the road safety authorities. This information is complete for 13 of the 15 regions, no data was obtained from Basque Country, and only a partial data set was obtained from Catalonia because both have a data collection system in place that is independent from the rest of the country. Data represented only those collisions that generated police report due to vehicle damages or motorist injuries. Road nomenclature and kilometer point was transformed in geographic coordinates (longitude and latitude) and then associated to a 10 × 10 km grid. WVC with a location error > 5 km were discarded. A total of 9508 WVC distributed in 2311 10 × 10 km cells were obtained for wild boar (see Supporting information), while up to 7029 WVC located in 1419 10 × 10 km cells were collected for roe deer.

Covariate predictors

Several bioclimatic and land cover covariates that usually affect ungulate abundance in Spain were obtained for each 10 × 10 km cell (Table 1). After accounting for multicollinearity in an original set of 51 predictors (ENETWILD Consortium et al. 2020), three bioclimatic variables (BIO) related with seasonality, temperature and precipitations were obtained from the Worldclim 2 project (<https://worldclim.org/version2>; Fick and Hijmans 2017). Percentage of relevant vegetation land cover type (LC) at each cell was calculated from the ESA/CCI-LC project, ver. v2.1.1 (2017) database (<www.esa-landcover-cci.org/?q=node/158>). A correlation plot of the final selected predictors for species abundance is shown in Supporting information. In addition, other predictors potentially involved in WVC were obtained. For each cell, kilometers of three kinds of roads were computed: highways (HWY), conventional roads (primary and secondary, ROAD) and urban (URB) roads (see Supporting information; National Geographic Institute, Ministry of Transport, Mobility and Urban Agenda 2021, <http://centrodedescargas.cnig.es/CentroDescargas/catalogo.do?Serie=CAANE>). Road km of each type at each cell were transformed into a three-level categorical factor (low, medium and high densities of each type of road). We used this classification as a simple proxy of average speed (HWY = 101.99 km h^{-1}, ROAD = 72.93 km h^{-1}, URB < 50 km h^{-1}; average for 2017) and traffic density (HWY = 25 584.17 vehicles per day (vh d^{-1}), ROAD = 4106 vh d^{-1}, URB = 9204 vh d^{-1}; average for 2017) since speed and traffic data for all roads in the whole study area were not available. Moreover, some authors have highlighted that the annual average daily traffic not always performs well as a predictor of WVC, since it does not account actual traffic density at the time when the collision took place (Bíl et al. 2020). Human influence index (HFP), obtained from The Last of the Wild Project

Table 1. Site specific covariates used to model the abundance and individual detectability of wild boar *Sus scrofa* and roe deer *Capreolus capreolus* based on wildlife–vehicle collisions data. Land cover variables represented percentage of cover per 10 × 10 km cells.

Code	Covariate description	Type
Abundance predictors		
BIO4	Temperature seasonality (temperature standard deviation × 100)	Continuous
BIO11	Mean temperature of coldest quarter	Continuous
BIO17	Precipitation of driest quarter	Continuous
LC10	Cropland, rainfed	Continuous
LC11	Herbaceous cover	Continuous
LC12	Tree or shrub cover	Continuous
LC20	Cropland, irrigated or post-flooding	Continuous
LC40	Mosaic natural vegetation (tree, shrub, herbaceous cover)/cropland	Continuous
LC60	Tree cover, broad-leaved, deciduous, closed to open	Continuous
LC70	Tree cover, needle leaved, evergreen, closed to open	Continuous
LC100	Mosaic tree and shrub/herbaceous cover	Continuous
LC120	Shrubland	Continuous
LC130	Grassland	Continuous
Individual detectability predictors		
HYW	Density of highways	Categorical
ROAD	Density of conventional roads	Categorical
URB	Density of urban roads	Categorical
MONTH	Month	Categorical
REG	Autonomous community (region)	Categorical
HFP	Human influence index	Continuous
TRI	Topographic ruggedness index	Continuous

ver. 2 (<http://sedac.ciesin.columbia.edu/data/collection/wildareas-v2>), was used as a human population density proxy. A topographic ruggedness index (TRI) was also computed for each cell from a digital elevation model (National Geographic Institute, Ministry of Transport, Mobility and Urban Agenda 2021, <https://www.ign.es/web/seccion-elevaciones>). Finally, each cell was assigned to the corresponding autonomous community (REG), since the WVC data availability was different for each region. Other specific factors related to the different regions can also affect to the WVC data, such as specific legislation about data reporting or policies about wildlife management like types of fences in hunting grounds (Gunson et al. 2011).

Royle and Nichols model (RN)

Single season site-occupancy models (MacKenzie et al. 2002) are a hierarchical modeling approach to model species occupancy (probability of a site is actually occupied, Ψ) that account for imperfect detection in response variable:

$$z_i \sim \text{Bernoulli}(\psi)$$

where z_i is the random variable presence or absence of species in site i. They relay in detection/non-detection data from repetitive K surveys at each site to model species detectability (MacKenzie et al. 2002):

$$h_{ij} \mid z_i \sim \text{Bernoulli}(z_i p_j)$$

where h_{ij} is the outcome of the jth survey in site i (detection/non-detection), z_i is the random variable presence or absence of species in site i and p_j is the detection probability for the jth survey. With this framework both presence probability **Ψ** and detection probability p can be modeled as function of site and survey specific covariates using any adequate link function (logit, etc.).

This framework was extended by Royle and Nichols (2003) to account for heterogeneity in detection probabilities as function of species abundance at each site. This model assumes a relationship between probability of detection p_i and abundance (N_i number of individuals in site i) in the way

$$N_i \sim \text{Poisson}(\lambda_i)$$

$$y_{ij} \mid N_i \sim \text{Bernoulli}(p_{ij}),$$

$$p_{ij} = 1 - (1 - r_{ij})^{Ni}$$

where y_{ij} is the detection outcome at unit i in the j survey and r_{ij} is the per-individual detection probability. Note that in our case the abundance is modeled through a Poisson distribution with λ parameter since we preferred to keep the modelling procedure as simple as possible and other distribution not always lead in easily interpretable results (for a discussion about unestable maximum likelihood estimations in RN models under negative binomial distribution see Royle and Nichols 2003), but this framework can accomplish any other discrete-valued distribution (negative-binomial, zero-inflated Poisson, etc.). Therefore, this model offers the possibility to estimate species abundance from detection/non-detection data registered in repetitive surveys. In addition, as in site-occupancy models, covariates affecting to abundance N and individual detectability r can be included by using link functions. Nevertheless, it is important to mention that N_i obtained for each site might be interpreted as a random effect yielding variation in p_i, not an abundance per se (MacKenzie et al. 2017). For this reason, N obtained from our RN models should be interpreted as an abundance index rather than actual individual abundance at each site (but see Linden et al. 2017).

Model fitting, selection and validation with hunting yield statistics

We used RN models (Royle and Nichols 2003) to obtain population abundance indices for wild boar and roe deer in mainland Spain from WVC data. We considered each month of 2017 as a survey occasion. For each 10 × 10 km cell (sites), a detection was annotated if at least one WVC was recorded at each month (survey). Otherwise, a non-detection was established. Site specific (bioclimatic, land cover, road km, etc.) and survey specific (month) covariates were used to fit a RN model accounting for heterogeneity in detectability driven by species abundance (Table 2). In a first step, we focused on modeling species abundance, using a general model for detection probability (all covariates included for individual detection process modeling). Firstly, we modeled abundance only as function of bioclimatic covariates (BIO), then only as function of land cover covariates (LC), and finally as function of both (BIO + LC). This strategy avoids constrains imposed in detection probability to focus on abundance as the ecological parameter of interest (MacKenzie et al. 2017). Once the abundance was fitted, we followed a similar procedure with the detectability, modeling it as function of road density and month first (HWY + ROADS + URB + MONTH), as function of road density, month, region and human influence (HWY + ROADS + URB + MONTH + REG + HFP) and finally as function of road density, month, region, human influence and topographic ruggedness index (HWY + ROADS + URB + MONTH + REG + HFP + TRI). All continuous variables were standardized before to perform the analyses.

We used the function *occuRN* from the R package *unmarked* to fit all models by maximum likelihood estimation (Fiske and Chandler 2011). Model selection based on Akaike's information criterion (AIC) was used to select the most appropriate model (Burnham and Anderson 2002). To

Table 2. Model selection results for abundance and individual detectability processes for wild boar *Sus scrofa* and roe deer *Capreolus capreolus* based on Akaike's information criterion. The number of parameters in the model (par) is also shown. Variables coded as Table 1.

Species	Model	par	AIC	ΔAIC
Abundance (λ_i)				
Wild boar	r_{ij}(HWY+ROAD+URB+MONTH+REG+HFP+TRI) λ_i(BIO4+BIO11+BIO17+LC10+LC11+LC12+LC20+LC40+LC60+LC70+LC100+LC120+LC130)	48	32 192.2	0
	r_{ij}(HWY+ROAD+URB+MONTH+REG+HFP+TRI) λ_i(LC10+LC11+LC12+LC20+LC40+LC60+LC70+LC100+LC120+LC130)	45	32 229.6	37.5
	r_{ij}(HWY+ROAD+URB+MONTH+REG+HFP+TRI) λ_i(BIO4+BIO11+BIO17)	38	32 329.1	136.8
Detectability (r_{ij})				
	r_{ij}(HWY+ROAD+URB+MONTH+REG+HFP+TRI) λ_i(BIO4+BIO11+BIO17+LC10+LC11+LC12+LC20+LC40+LC60+LC70+LC100+LC120+LC130)	48	32 192.2	0
	r_{ij}(HWY+ROAD+URB+MONTH+REG+HFP) λ_i(BIO4+BIO11+BIO17+LC10+LC11+LC12+LC20+LC40+LC60+LC70+LC100+LC120+LC130)	47	32 232.8	40.6
	r_{ij}(HWY+ROAD+URB+MONTH) λ_i(BIO4+BIO11+BIO17+LC10+LC11+LC12+LC20+LC40+LC60+LC70+LC100+LC120+LC130)	32	33 210.5	1018.3
Abundance (λ_i)				
Roe deer	r_{ij}(HWY+ROAD+URB+MONTH+REG+HFP+TRI) λ_i(BIO4+BIO11+BIO17+LC10+LC11+LC12+LC20+LC40+LC60+LC70+LC100+LC120+LC130)	48	21 469.8	0
	r_{ij}(HWY+ROAD+URB+MONTH+REG+HFP+TRI) λ_i(BIO4+BIO11+BIO17)	38	21 666.4	196.6
	r_{ij}(HWY+ROAD+URB+MONTH+REG+HFP+TRI) λ_i(LC10+LC11+LC12+LC20+LC40+LC60+LC70+LC100+LC120+LC130)	45	21 823.1	353.2
Detectability (r_{ij})				
	r_{ij}(HWY+ROAD+URB+MONTH+REG+HFP+TRI) λ_i(BIO4+BIO11+BIO17+LC10+LC11+LC12+LC20+LC40+LC60+LC70+LC100+LC120+LC130)	48	21 469.8	0
	r_{ij}(HWY+ROAD+URB+MONTH+REG+HFP) λ_i(BIO4+BIO11+BIO17+LC10+LC11+LC12+LC20+LC40+LC60+LC70+LC100+LC120+LC130)	47	21 625.8	156.1
	r_{ij}(HWY+ROAD+URB+MONTH) λ_i(BIO4+BIO11+BIO17+LC10+LC11+LC12+LC20+LC40+LC60+LC70+LC100+LC120+LC130)	32	22 545.1	1075.3

assess the goodness-of-fit for the RN model we used parametric bootstrapping (MacKenzie and Bailey 2004) and calculated the overdispersion parameter (ĉ) using the function *mb.gof.test* from the R package *AICcmodavg* (Mazerolle 2020). We generated spatial predictions (abundance and detectability) for each species for the grid cells in the whole study area. It is important to mention that one of the assumptions of RN models is that animal populations are geographically and demographically closed, that is, number of animals at each cell does not change overtime within the season, which means no deaths, births, immigration or emigration. Although this is presumably not true in our study, we assume that RN models are robust to slight variations in population size as long as occupancy state does not change, which is mostly true in our case.

To assess abundance index values obtained from WVC data for each species, we aggregated cell predictions at province level for mainland Spain and compared it with raw hunting yields obtained for 2017 season using a least-squares regression (total number of hunted animal per each region, Ministry of Agriculture, Food and Environment published 'Yearbooks of Forestry Statistics' (<www.mapa.gob.es/es/desarrollo-rural/estadisticas/forestal_anuarios_todos.aspx>), compiled from the reports submitted by the provincial hunting departments).

Results

Model fitting and selection

The results for model selection using AIC are shown in Table 2. For the two species, the first stage in the fitting procedure indicated that variation in abundance index was driven by both, bioclimatic and land cover predictors. For wild boar, model with only land cover covariates obtained a lower AIC than the model with only bioclimatic predictors, while the opposite pattern was obtained for roe deer (Table 2). The second step focused on individual detectability showed similar results for both species, being the best model the one with all covariates (Table 2). Therefore, we selected the saturated model in both processes to predict wild boar and roe deer abundance index across mainland Spain.

All model estimates and their standard errors for the abundance and individual detectability for both species are shown in the Supporting information. In the abundance process from wild boar model, confident intervals of estimates for BIO11, LC11, LC12, LC100 and LC120 overlapped with 0, which indicated a low effect of these covariates (Fig. 1). The rest of covariates showed a positive relationship with the abundance process (Fig. 1). For roe deer, none of the estimate's confident intervals overlapped with 0. Bioclimatic

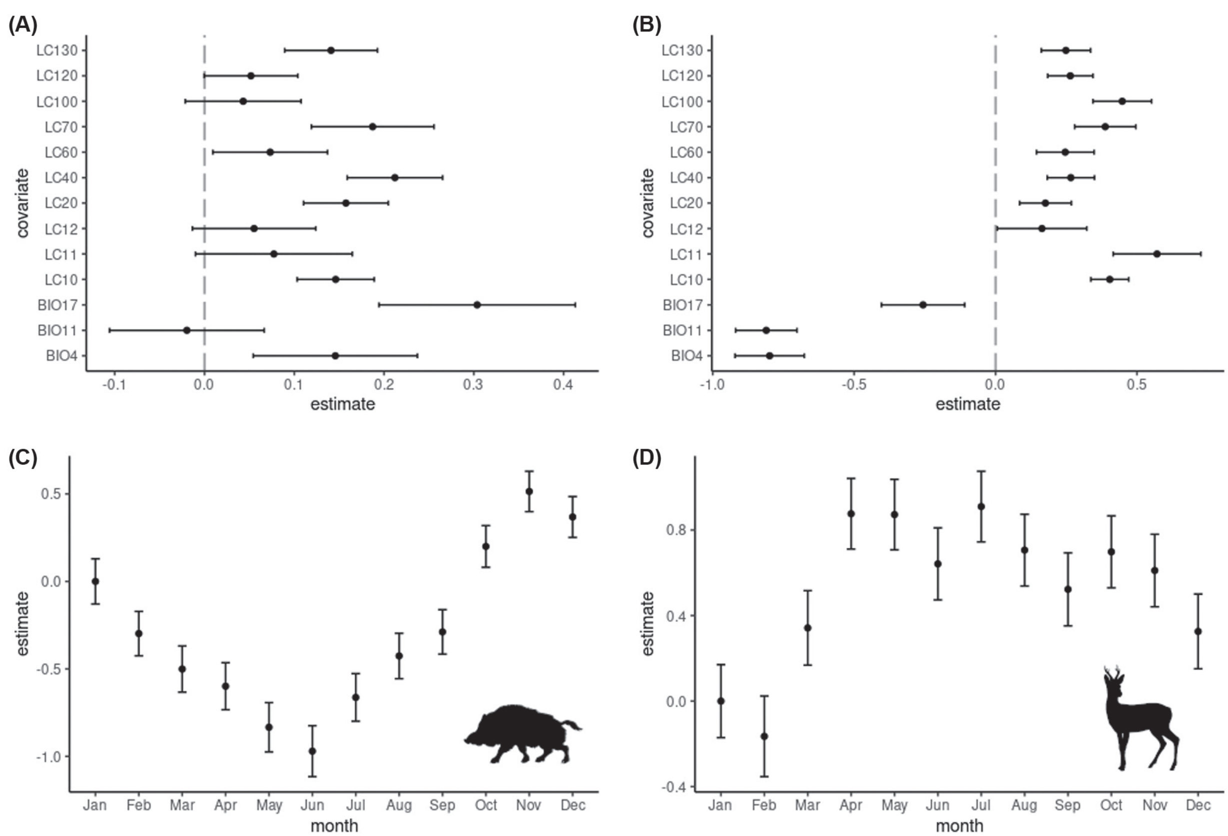

Figure 1. Results for full models (including all covariates in both processes) for wild boar (A, C) and roe deer (B, D). Covariate estimates for abundance process (A, B) and effect of the survey specific covariate (month) in individual detectability process (C, D). Bioclimatic and land cover variables are defined in Table 1. Bars represent standard errors of estimates.

variables BIO4, BIO11 and BIO17 showed a negative relation with roe deer abundance process, while all the land cover variables showed a positive relationship (Fig. 1). Detectability of both species highly varied among months, being October, November and December those with higher detectability for wild boar, while April–July was the period with higher detectability for roe deer (Fig. 1). For wild boar the three types of roads (highways, conventional roads and urban roads) increased the detectability at higher road densities. Similar results were obtained for roe deer excepting for highways, which showed a non-linear effect decreasing detectability at medium highways densities (see Supporting information). Two regions showed especially low detectability values for wild boar, Basque Country and Catalonia, while Murcia Region obtained the lowest detectability for roe deer. Human index (positive) and topographic ruggedness (negative) showed a similar effect over individual detectability for both species. The goodness-of-fit statistics indicated non overdispersion for wild boar ($\hat{c} = 1.09$), nor for roe deer ($\hat{c} = 0.99$).

Model predictions and comparison with hunting yield statistics

Model predictions for wild boar and roe deer abundance and individual detectability are shown in Fig. 2. To spatially predict individual detectability, both models were fitted to January for the month covariate. Spatial pattern of wild boar abundance index described a northeast–southwest gradient with the highest abundances in Pyrenean and pre-Pyrenean Mountains (see Supporting information, provinces 32, 9, 29 and 28). Wild boar was also abundant in the rest of North Spain, with a significant decrease for the interior of northern plateau (Castilla y León region; Fig. 2A). Roe deer was abundant in Cantabrian, Iberian and Pyrenean Mountains, while the southern Spain obtained the lowest abundances

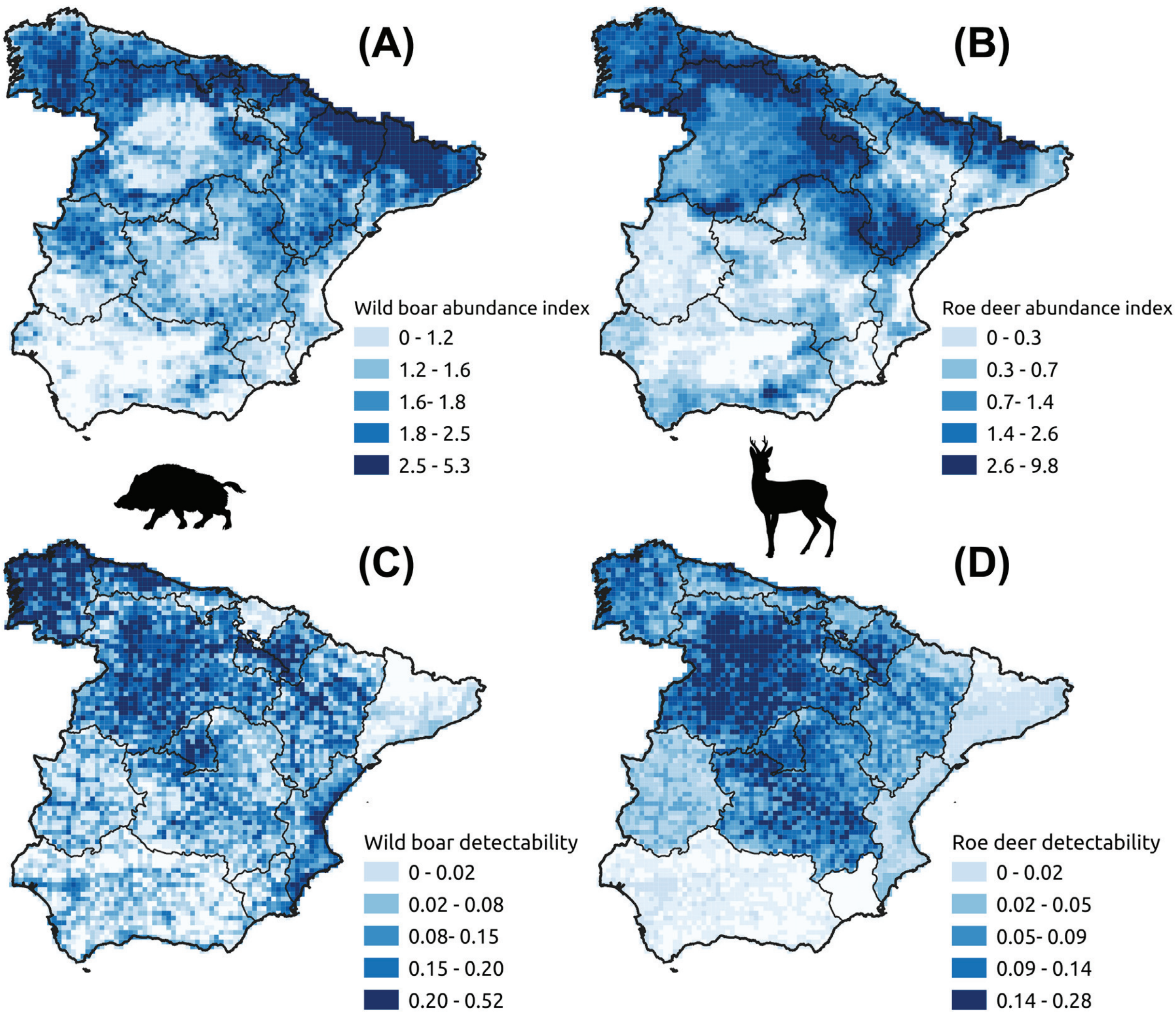

Figure 2. Results for full models (including all covariates in both processes) for wild boar (A, C) and roe deer (B, D). Abundance index (A, B) and individual detectability (C, D) predictions for wild boar and roe deer respectively. Note that individual detectability was fixed for month = January.

(Fig. 2B). Individual detectability showed a scattered pattern for both species due to differences in road densities at each 10 × 10 km cell. In general, the northern half of mainland Spain obtained higher detectability, especially for the regions of Galicia and Castilla y León (Fig. 2C and D).

Least square regression results and correlation plots between model predictions and raw hunting yields at province level are shown in Fig. 3. Both abundance models showed a positive correlation with raw hunting yields (wild boar R = 0.78; n = 47 and $p < 0.01$; roe deer R = 0.63, n = 47 and $p < 0.01$), but the relationship was stronger for wild boar than for roe deer.

Discussion

Accurate information about wildlife population abundance is necessary for management and/or conservation plans. However, for broadly distributed or non-emblematic generalist species, this information is not always easy to obtain at national or regional scale due to the important sampling effort needed to cover the whole study area. Thus, the search for different and complementary sources of wildlife abundance information and the development of methodological frameworks to deal with them is a key stage in broad scale monitoring programs (ENETWILD Consortium et al. 2018). Wildlife vehicle collisions offer the possibility to study not only factors affecting road casualties (Gunson et al. 2011), but also the drivers of wildlife abundance if the ecological process is disentangled from the collision (detection) process. We presented empirical evidence that, when individual detectability is considered in the modeling process, wildlife abundance indices can be generated from WVC data with similar performance than other commonly used at large spatial scale abundance indices, such as hunting statistics (Imperio et al. 2010). Although similar methods have been used to study animal roadkill occurrence (Santos et al. 2018), as far as we know this is the first time using Royle–Nichols models to derive large scale species abundance indices from WVC data.

Royle–Nichols model for wildlife population abundance index modeling from WVC data

Our results suggest that the Royle–Nichols modeling approach produce accurate indices of wildlife population abundance from WVC data. It is important to note the effect of accounting for detectability in the predicted abundance pattern from WVC data (Supporting information and Fig. 3A). In the case of wild boar, we can observe that higher abundances not always correspond to those areas with higher WVC (interior of northern plateau, Castilla y León region; Fig. 2A). Higher WVC rates can be due to higher individual detectability (Fig. 2C) and therefore, the lack of control of this process could led in misleading results when using directly WVC as wildlife abundance index (Rodríguez-Morales et al. 2013). Similarly, the absence of WVC is not always produced by low population abundance index. Our approach correctly detected lower detectability in some regions where WVC were underrepresented in our database (Basque Country and Catalonia for wild boar, south of Andalusia for roe deer), and predicted accurate abundance indices for those regions (Fig. 3).

Although other authors have reached accurate absolute abundance estimations using RN models (Linden et al. 2017), only relative abundance indices were obtained in our study case. Royle–Nichols model simply accommodates heterogeneity in detection probability that is formally attributed to variation in abundance, but other sources of heterogeneity could be affecting our data (MacKenzie et al. 2017). Absolute abundance estimates from our model predictions are far to reach the actual population densities estimated by other more

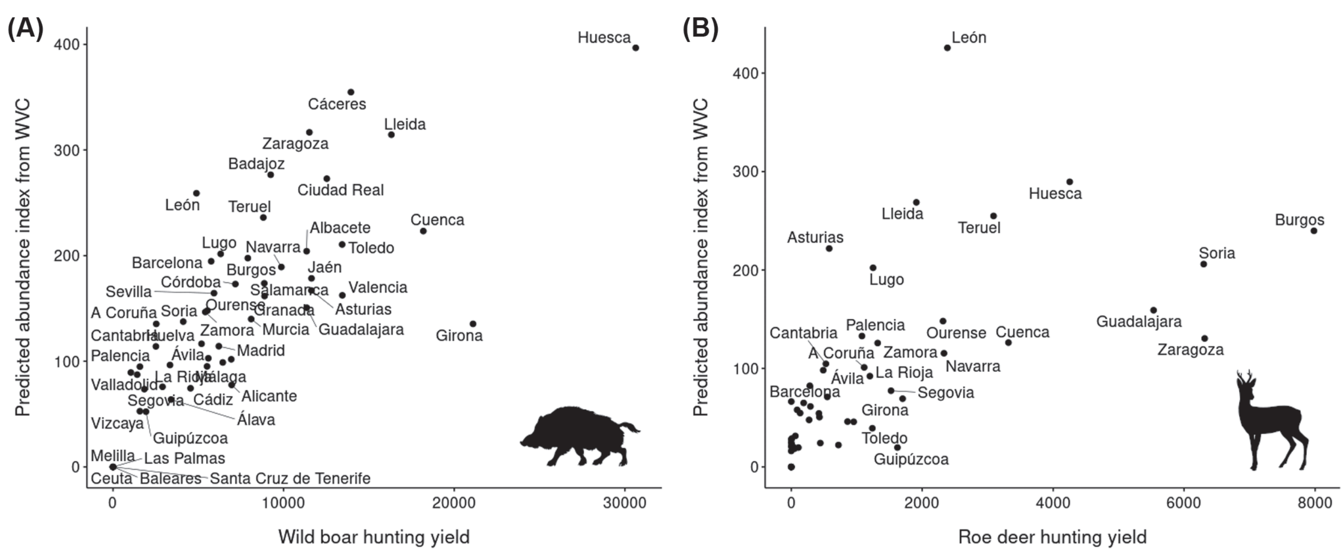

Figure 3. Relationship between abundance index derived from full models using wildlife vehicle collision data and raw hunting yields (number of hunted animals) at province level for mainland Spain in 2017 (A, wild boar R = 0.78, $R^2 = 0.61$; B, roe deer R = 0.63, $R^2 = 0.4$).

specific methods (hunting yields, telemetry, camera trap, etc.) and therefore our predictions should be interpreted as abundance indices. Despite this, the potential of RN model to predict abundance indices from WVC opens a promising field that in combination to other data sources will lead in more accurate absolute abundance estimates.

In addition to population abundance, RN model allow us to obtain spatial and temporal predictions on detectability, which can be used as a wildlife vehicle collision risk index to detect black spots for animal and motorist casualties in roads (van der Grift et al. 2013). However, this is conditioned to the systematic register of WVC in our databases, since low detectability can be also due to low rate of WVC report by administrations or incomplete databases as in our case (Basque Country and Catalonia).

Factors affecting wild boar and roe deer vehicle collisions

One of the advantages of the RN modeling approach is the possibility to disentangle abundance and individual detectability processes. The most important covariates affecting wild boar abundance were precipitation of driest quarter, mosaic natural vegetation/cropland and evergreen tree cover (Fig. 2, Supporting information). This is in accordance with previous studies about wild boar abundance in Mediterranean regions, where water and therefore resource availability during the driest season strongly affects wild boar and other ungulates abundances (ENETWILD Consortium et al. 2020). Similarly, mosaic or transition areas between natural vegetation and croplands usually provide both refugee and feeding resources and have been already related to higher wild boar abundances (Acevedo et al. 2014). Seasonality in both temperature and precipitation were negatively related to roe deer abundance (Fig. 2). It has been already described the relationship between roe deer distribution and Atlantic influenced areas in the Iberian Peninsula, in which water availability is more constant across the year and temperatures are milder (Aragón et al. 1995, Acevedo et al. 2005). On the contrary, croplands-herbaceous and patched landscape with tree or shrub and herbaceous cover were related to higher roe deer abundances. This cervid has experienced a range expansion during the last decades in the Iberian Peninsula from forest areas (Virgós and Tellería 1998). It has promoted the colonization of heterogeneous habitats and grasslands previously used in agriculture which could explain higher abundances related to croplands and herbaceous covers (Acevedo et al. 2005).

Our modeling approach was able to control the lack of data for some regions (namely Basque Country and Catalonia) which is represented by the low detectability at those regions for both species (Supporting information). We obtained a low detectability for roe deer in Murcia region, in which it is certainly absent for the most of the territory. We also captured the temporal variation in individual detectability for wild boar and roe deer. While autumn was the season with higher probabilities for wild boar vehicle collisions, roe deer causalities were more likely in spring and early summer (Fig. 2). This pattern has been previously reported and could be due to several factors apart from traffic variation (Jacobson et al. 2016). Collective hunting activities during October–December in Spain can increase the wild boar movements (Maillard and Fournier 1995), which can lead in higher frequency of road-crossing and therefore higher collision probabilities. However, other confounding factors such as longer nights in these months or reproductive behavior could also play a role (Lagos et al. 2012). In the case of roe deer, Lagos et al. (2012) related higher collision probabilities in April–June and July to the mother–fawn separation prior to the calving and to the rutting season, respectively. Like wild boar, high animal mobility in these periods could increase the probability of WVC. The rest of covariate effects were expected, included the decrease of detectability at higher densities of highways for roe deer (Supporting information), since many authors have described roe deer roads avoidance (Madsen et al. 2002, Torres et al. 2011, Kušta et al. 2017).

Hunting yields versus wildlife vehicle collision data as abundance indices

Hunting yields statistics are used for game species abundance monitoring at regional and broader spatial scales (ENETWILD Consortium et al. 2019). Thus, HY were used in our study as validation dataset to assess abundance index obtained from WVC. Positive linear correlations indicated similar outcomes when comparing raw HY with model abundance predictions from WVC for both species (Fig. 3). We found stronger relation in wild boar than roe deer, which could be related to the broader distributional range of wild boar compared to roe deer. Nevertheless, those relationships have been already observed for other ungulates in Europe (Wiebke et al. 2020) or North America (McCaffery 1973). This validation process should be replicated for other countries to empirically assess the robustness of wildlife abundance indices derived from WVC data (Ascensão et al. 2019b). Wildlife monitoring programs at large scales require systematic sampling effort across different regions or countries. Establishing coordinated networks of collaborators to apply specific sampling methods such as camera tramp or direct counts across different countries is usually unaffordable due to cost expenses (Burton et al. 2015). For this reason, hunting statistics in combination with another calibrated method such as drive counts have been proposed as a low-cost methodology to infer species abundance at regional or boarder scales, offering the possibility to provide long term trends in population dynamics (ENETWILD Consortium et al. 2018). However, not all species, nor all regions can benefit from these approaches. In this study, we illustrated that RN models using detection/non-detection data from WVC produce similar inferences about relative abundance to hunting statistics in mainland Spain. Wildlife vehicle collision data also share several useful features for large scale monitoring programs: data is systematically registered by administrations or other organizations in countries with a well-developed primary and secondary road systems, the volume of data is often enormous among large species and historical or long time series can be easily obtained (Romin and Bissonette 1996, Brockie et al. 2009). Detection/non-detection models that account for imperfect detectability can control those factors that could affect abundance estimates from WVC data such as traffic density or vehicles speed. At the same time, they provide spatially and temporal explicit predictions about individual detectability, that is, wildlife vehicle collision risk, which can be used in mitigation programs to avoid animal and motorist mortality.

Concluding remarks and recommendations

The Royle–Nichols model applied to WVC data can identify the main drivers of the wild boar and roe deer abundance in mainland Spain (ecological process) as well as those factors affects the animal vehicle collision probability (observational process). Abundance indices obtained from WVC are like those obtained from the hunting yield statistics, one of the most important data sources for abundance monitoring in game species. Our approach provides new opportunities in the study of wildlife abundance distribution for those territories where hunting bag information is not accurate enough or such data is not available. In addition, time series of WVC data could be used for monitoring population trends in a multi season framework. In this context, abundance predictions from the Royle–Nichols models should be evaluated to study their ability to be extrapolated, down or upscaled, etc., using other independent sources such as telemetry or camera trapping studies (Palencia et al. 2021).

Acknowledgements – We thank Dirección General de Tráfico (DGT) of the Spanish Ministry of the Interior for data collection and supply. Sonia Illanas and Pablo Palencia helped with overdispersion parameter computation through Servicio de Supercomputacion from Univ. de Castilla – La Mancha.

Funding – This research was funded by the Spanish Ministry of Science (Project: 'Armonización de los datos poblacionales de la fauna silvestre en España: aplicaciones a la vigilancia sanitaria y control de enfermedades compartidas con el ganado.' Ref. PID2019-111699RB-I00, HAWIPO). JFL was partially funded by the grant NSF – 1759940, and currently by the grant Margarita Salas from the European Union – NextGenerationEU through the Complutense University.

Author contributions

Javier Fernández-López: Conceptualization (lead); Data curation (lead); Formal analysis (lead); Writing – original draft (lead); Writing – review and editing (lead). **José A. Blanco-Aguiar**: Conceptualization (equal); Data curation (equal); Resources (equal); Writing – review and editing (equal). **Joaquín Vicente**: Conceptualization (equal); Data curation (equal); Resources (equal); Writing – review and editing (equal). **Pelayo Acevedo**: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Supervision (lead); Writing – review and editing (equal).

Transparent peer review

The peer review history for this article is available at <https://publons.com/publon/10.1111/ecog.06113>.

Data availability statement

The dataset used in this study contains sensitive information and is available upon request from the Spanish General Traffic Agency (Dirección General de Tráfico - DGT). Requests regarding the dataset should be made to servicio.estadistica@dgt.es.

Supporting information

The supporting information associated with this article is available from the online version.

References

Acevedo, P. et al. 2005. Environmental constraints in the colonization sequence of roe deer (*Capreolus capreolus* Linnaeus, 1758) across the Iberian Mountains, Spain. – J. Biogeogr. 32: 1671–1680.
Acevedo, P. et al. 2011. Past, present and future of wild ungulates in relation to changes in land use. – Landsc. Ecol. 26: 19–31.
Acevedo, P. et al. 2014. Spatial distribution of wild boar population abundance: basic information for spatial epidemiology and wildlife management. – Ecol. Indic. 36: 594–600.
Aragón, S. et al. 1995. Socioeconomic, physiognomic and climatic factors determining the distribution pattern of roe deer *Capreolus capreolus* in Spain. – Acta Theriol. 40: 37–43.
Ascensão, F. et al. 2019a. Beware that the lack of wildlife mortality records can mask a serious impact of linear infrastructures. – Global Ecol. Conserv. 19: e00661.
Ascensão, F. et al. 2019b. Validation data is needed to support modelling in Road Ecology. – Biol. Conserv. 230: 199–200.
Ascensão, F. et al. 2021. Wildlife collisions put a dent in road safety. – Science 374: 1208–1208.
Baker, P. J. et al. 2004. Is it possible to monitor mammal population changes from counts of road traffic casualties? An analysis using Bristol's red foxes *Vulpes vulpes* as an example. – Mammal Rev. 34: 115–130.
Barrientos, R. and Miranda, J. D. 2012. Can we explain regional abundance and road-kill patterns with variables derived from local-scale road-kill models? Evaluating transferability with the European polecat. – Divers. Distrib. 18: 635–647.
Barrientos, R. et al. 2021. The lost road: do transportation networks imperil wildlife population persistence? – Perspect. Ecol. Conserv. 19: 411–416.
Bíl, M. et al. 2020. Ungulate-vehicle collision risk and traffic volume on roads. – Eur. J. Wildl. Res. 66: 59.
Bissonette, J. A. et al. 2008. Assessment of costs associated with deer–vehicle collisions: human death and injury, vehicle damage and deer loss. – Hum.-Wildl. Confl. 2: 17–27.
Brockie, R. E. et al. 2009. Long-term wildlife road-kill counts in New Zealand. – New Zeal. J. Zool. 36: 123–134.

Burnham, K. P. and Anderson, D. R. 2002. Model selection and multimodel inference: a practical information–theoretic approach. – Springer.
Burton, A. C. et al. 2015. Wildlife camera trapping: a review and recommendations for linking surveys to ecological processes. – J. Appl. Ecol. 52: 675–685.
Colino-Rabanal, V. J. and Peris, S. J. 2016. Wildlife roadkills: improving knowledge about ungulate distributions? – Hystrix 27: 91–98.
D'Amico, M. et al. 2015. Vertebrate road-kill patterns in Mediterranean habitats: who, when and where. – Biol. Conserv. 191: 234–242.
ENETWILD consortium et al. 2018. Guidance on estimation of wild boar population abundance and density: methods, challenges, possibilities. – EFSA Support. Publ. 15: 1449E.
ENETWILD consortium et al. 2019. Harmonization of the use of hunting statistics for wild boar density estimation in different study areas: report based on comparison of case studies in different wild boar populations representative of the different management and habitat conditions across Europe. – EFSA Support. Publ. 16: 1706E.
ENETWILD consortium et al. 2020. Update of occurrence and hunting yield-based data models for wild boar at European scale: new approach to handle the bioregion effect. – EFSA Support. Publ. 17: 1871E.
Fick, S. E. and Hijmans, R. J. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. – Int. J. Climatol. 37: 4302–4315.
Fiske, I. and Chandler, R. 2011. Unmarked: an R package for fitting hierarchical models of wildlife occurrence and abundance. – J. Stat. Softw. 43: 1–23.
Garrote, G. et al. 2018. Prediction of Iberian lynx road–mortality in southern Spain: a new approach using the MaxEnt algorithm. – Anim. Biodivers. Conserv. 41: 217–225.
Grilo, C. et al. 2020. Roadkill risk and population vulnerability in European birds and mammals. – Front. Ecol. Environ. 18: 323–328.
Gunson, K. E. et al. 2011. Spatial wildlife–vehicle collision models: a review of current work and its application to transportation mitigation projects. – J. Environ. Manage. 92: 1074–1082.
Imperio, S. et al 2010. Investigating population dynamics in ungulates: do hunting statistics make up a good index of population abundance? – Wildl. Biol. 16: 205–214.
Jacobson, S. L. et al. 2016. A behavior-based framework for assessing barrier effects to wildlife from vehicle traffic volume. – Ecosphere 7: e01345.
Jones, J. P. 2011. Monitoring species abundance and distribution at the landscape scale. – J. Appl. Ecol. 48: 9–13.
Kušta, T. et al. 2017. The effect of traffic intensity and animal activity on probability of ungulate-vehicle collisions in the Czech Republic. – Safety Sci. 91: 105–113.
Lagos, L. et al. 2012. Temporal pattern of wild ungulate-related traffic accidents in northwest Spain. – Eur. J. Wildl. Res. 58: 661–668.
Linden, D. W. et al. 2017. Examining the occupancy–density relationship for a low-density carnivore. – J. Appl. Ecol. 54: 2043–2052.
MacKenzie, D. I. and Bailey, L. L. 2004. Assessing the fit of site-occupancy models. – J. Agric. Biol. Environ. Stat. 9: 300–318.
MacKenzie, D. I. et al. 2002. Estimating site occupancy rates when detection probabilities are less than one. – Ecology 83: 2248–2255.
MacKenzie, D. I. et al. 2017. Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence. – Elsevier.
Madsen, A. B. et al. 2002. Factors causing traffic killings of roe deer *Capreolus capreolus* in Denmark. – Wildl. Biol. 8: 55–61.
Maillard, D. and Fournier, P. 1995. Effects of shooting with hounds on size of resting range of wild boar (*Sus scrofa* L.) groups in Mediterranean habitat. – J. Mount. Ecol. 3: 102–107.
Mallick, S. A. et al. 1998. Road-kills of the eastern barred bandicoot *Perameles gunnii* in Tasmania: an index of abundance. – Wildl. Res. 25: 139–145.
Massei, G. et al. 2015. Wild boar populations up, numbers of hunters down? A review of trends and implications for Europe. – Pest Manage. Sci. 71: 492–500.
Mazerolle, M. J. 2020. AICcmodavg: model selection and multimodel inference based on (Q)AIC(c). R package ver. 2.3-1. – <https://cran.r-project.org/package=AICcmodavg>.
McCaffery, K. R. 1973. Road-kills show trends in Wisconsin deer populations. – J. Wildl. Manage. 37: 212–216.
Palencia, P. et al. 2021. Assessing the camera trap methodologies used to estimate density of unmarked populations. – J. Appl. Ecol. 58: 1583–1592.
Pearce, J. L. and Boyce, M. S. 2006. Modelling distribution and abundance with presence-only data. – J. Appl. Ecol. 43: 405–412.
Rodríguez-Morales, B. et al. 2013. Spatiotemporal analysis of vehicle collisions involving wild boar and roe deer in NW Spain. – Accid. Anal. Prevent. 60: 121–133.
Romin, L. A. and Bissonette, J. A. 1996. Deer: vehicle collisions: status of state monitoring activities and mitigation efforts. – Wildl. Soc. Bull. 24: 276–283.
Royle, J. A. and Nichols, J. D. 2003. Estimating abundance from repeated presence–absence data or point counts. – Ecology 84: 777–790.
Rytwinski, T. and Fahrig, L. 2012. Do species life history traits explain population responses to roads? A meta-analysis. – Biol. Conserv. 147: 87–98.
Sáenz-de-Santa-María, A. and Tellería, J. L. 2015. Wildlife–vehicle collisions in Spain. – Eur. J. Wildl. Res. 61: 399–406.
Saint-Andrieux, C. et al. 2020. Comparison of environmental, biological and anthropogenic causes of wildlife–vehicle collisions among three large herbivore species. – Popul. Ecol. 62: 64–79.
Santos, R. A. et al. 2018. Predicting wildlife road-crossing probability from roadkill data using occupancy-detection models. – Sci. Total Environ. 642: 629–637.
Torres, R. T. et al. 2011. Factors affecting roe deer occurrence in a Mediterranean landscape, Northeastern Portugal. – Mammal. Biol. 76: 491–497.
van der Grift, E. A. et al. 2013. Evaluating the effectiveness of road mitigation measures. – Biodivers. Conserv. 22: 425–448.
Vanlaar, W. G. et al. 2012. Wildlife vehicle collisions in Canada: a review of the literature and a compendium of existing data sources. – Traffic Injury Research Foundation and Eco-Kare International, 69 pp.
Vicente, J. et al. 2019. Science-based wildlife disease response. – Science 364: 943–944.
Virgós, E. and Tellería, J. L. 1998. Roe deer habitat selection in Spain: constraints on the distribution of a species. – Can. J. Zool. 76: 1294–1299.
Wiebke, N. et al. 2020. Strength of correlation between wildlife collision data and hunting bags varies among ungulate species and with management scale. – Eur. J. Wildl. Res. 66: 86.