ECOGRAPHY

Research article

Bayesian species distribution models integrate presence-only and presence–absence data to predict deer distribution and relative abundance

Virginia Morera-Pujol [1], Philip S. Mostert[2], Kilian J. Murphy[1], Tim Burkitt[3], Barry Coad[4], Barry J. McMahon[5], Maarten Nieuwenhuis[6], Kevin Morelle[7,8], Alastair I. Ward[9] and Simone Ciuti[1]

[1]Laboratory of Wildlife Ecology and Behaviour, School of Biology and Environmental Science, Univ. College Dublin, Dublin, Ireland
[2]Dept of Mathematical Sciences and Center for Biodiversity Dynamics, Norwegian Univ. of Science and Technology (NTNU), Trondheim, Norway
[3]Deer Biologist and Management Consultant, Deer Management Solutions, Coolies, Muckross, Killarney, Co. Kerry, Ireland
[4]Coillte Forest, Coillte, Dublin Road, Newtownmountkennedy, Co Wicklow, Ireland
[5]UCD School of Agriculture & Food Science, Univ. College Dublin, Belfield, Dublin, Ireland
[6]UCD Forestry, School of Agriculture and Food Science, Univ. College Dublin, Belfield, Dublin, Ireland
[7]Dept of Migration, Max Planck Inst. of Animal Behaviour, Radolfzell, Germany
[8]Dept of Game Management and Wildlife Biology, Czech Univ. of Life Science, Prague, Czech Republic
[9]School of Biology, Univ. of Leeds, Leeds, UK

Correspondence: Virginia Morera-Pujol (morera.virginia@gmail.com)

Using geospatial data of wildlife presence to predict a species distribution across a geographic area is among the most common tools in management and conservation. The collection of high-quality presence–absence (PA) data through structured surveys is, however, expensive, and managers usually have access to larger amounts of low-quality presence-only (PO) data collected by citizen scientists, opportunistic observations and culling returns for game species. Integrated species distribution models (ISDMs) have been developed to make the most of the data available by combining the higher-quality, but usually scarcer and more spatially restricted, PA data with the lower-quality, unstructured, but usually more extensive PO datasets. Joint-likelihood ISDMs can be run in a Bayesian context using integrated nested laplace approximation methods that allow the addition of a spatially structured random effect to account for data spatial autocorrelation. Here, we apply this innovative approach to fit ISDMs to empirical data, using PA and PO data for the three prevalent deer species in Ireland: red, fallow and sika deer. We collated all deer data available for the past 15 years and fitted models predicting distribution and relative abundance at a 25 km² resolution across the island. Model predictions were associated to spatial estimate of uncertainty, allowing us to assess the quality of the model and the effect that data scarcity has on the certainty of predictions. Furthermore, we checked the performance of the three species-specific models using two datasets, independent deer hunting returns and deer densities based on faecal pellet counts. Our work clearly demonstrates the applicability of spatially explicit ISDMs to empirical data in a Bayesian context, providing a blueprint for managers to exploit unexplored and seemingly unusable data that can, when modelled with the proper tools, serve to inform management and conservation policies.

Introduction

Methods to accurately predict species distributions have been central to wildlife management, conservation of endangered species, control of invasive species and improvement of human–wildlife coexistence (Nyhus 2016, Frans et al. 2021). Species distribution models (SDMs) correlate species occurrence to variables reflecting climatic and environmental conditions, allowing us to understand spatiotemporal drivers of species occurrence in different areas or under different climatic conditions (Guisan and Zimmermann 2000). SDMs have increased in complexity since their origin, aiming to improve the predictions based on environmental variables, to account for spatial autocorrelation and to include different data types such as presence-only (PO) or presence–absence (PA) (Guisan and Thuiller 2005, Elith and Leathwick 2009, Guillera-Arroita et al. 2015).

Most SDM methods require information on where the species has been recorded, where it has been absent from and, to avoid biases, require systematically collected data with strict control for effort, methodology and spatial coverage. These structured data, however, are typically expensive to collect and are thus scarce and with low spatial coverage (Hortal and Lobo 2005, Miller et al. 2019). Unstructured data, where collection effort and exact location may not be specified, sometimes even collected without a standardised protocol, are usually a more abundant but less accurate source of information with the potential to give relevant insights about species ecology. Unstructured data may range between museum records and opportunistic citizen science observations, sometimes collected using recent advances in technology such as smartphone applications (Boyce and Corrigan 2017, Pacifici et al. 2017); in game species, unstructured, low spatial resolution data can originate from culling returns (Nagy-Reis et al. 2021). Although unstructured datasets may be more abundant and have wider spatial and temporal coverage than structured data, their use in SDMs raises issues, such as the need to carefully consider observation bias as well as the underestimation of local occurrence rates due to the lack of information on the observational process (Yackulic et al. 2013, Pacifici et al. 2017).

In contrast to structured data, opportunistic datasets do not include species absences, and thus methods have been developed to work with PO data by generating or estimating pseudo-absences (i.e. locations where the species could have been present but were not observed; Lobo et al. 2010, Elith et al. 2011, Carlson 2020). Nevertheless, both data types are often available for a single species, area and time period, and the possibility of combining different datasets can allow for better inference across much larger spatial extents. Among the approaches developed to cope with this analytical challenge, two have gained traction in ecology studies: data pooling and model-based data integration (or integrated species distribution models, ISDMs; Fletcher et al. 2019, Isaac et al. 2020).

The data pooling approach combines datasets prior to entering a model, by degrading the higher-quality dataset until it has a common observation process with the lower-quality dataset (e.g. converting a PA dataset to PO observations, Ahmad Suhaimi et al. 2021). Alternatively, ISDMs avoid degrading the quality of the data in the most accurate dataset by considering the two datasets as different representations of the same distribution, and thus modelling them together, combining the two likelihoods (joint-likelihood approach, Pacifici et al. 2017). Additional advantages have become obvious in ISDMs: on the one hand, including an unbiased structured dataset (i.e. a PA dataset) helps compensate for potential biases in PO datasets (Dorazio 2014); on the other hand, ISDMs improve the ability to predict over a wider geographic area by combining a very often spatially restricted PA dataset with an overlapping, but more extensive, PO dataset (Ahmad Suhaimi et al. 2021).

As datasets become increasingly complex, the challenge for SDMs is to find appropriate ways to account for the spatial structure of the observations and their intrinsic autocorrelation. Several approaches have been adopted so far, including conditional autoregressive (CAR) models fitted within a Bayesian context through Markov chain Montecarlo (MCMC) methods (Zulian et al. 2021). Most recently, hierarchical Bayesian models fitted using integrated nested laplace approximation (INLA, Lindgren and Rue 2015) provide several advantages: on the one hand, they allow for the inclusion of a spatially structured random effect (i.e. spatial field) that can capture more complex spatial structures that might influence the distribution of observations; on the other hand, INLA allows for the modelling in continuous space (without the need to group the observations by grid cell) through the use of stochastic partial differential equations (SPDE, Lindgren et al. 2011, Paradinas et al. 2017, Lezama-Ochoa et al. 2020). This type of model can be implemented with the 'R-INLA' package (Rue et al. 2009, Bakka et al. 2018) and the more recent development of the 'inlabru' package (which provides easy access to most 'R-INLA' functionality for spatially structured data; Bachl et al. 2019). This approach provides a computationally fast modelling environment for hierarchical Bayesian models, where complex spatially structured random effects can be added to models for a wide variety of response variables (e.g. binomial models for PA data or Poisson models for PO and count data, Bakka et al. 2018), and where joint-likelihood models can be developed to integrate different sources of data (Adde et al. 2021, Cunningham et al. 2021, Simmonds et al. 2020).

These recent methodological approaches can boost the ability to model species distributions and disentangle the factors that drive habitat selection, thus making them a great tool for wildlife conservation and management (Linnell and Zachos 2011). This is particularly important when it comes to quickly assessing the status of species of

conservation concerns or, otherwise, overabundant species raising management challenges. Since the latter half of the 20th century, ungulate populations across Europe have increased rapidly, leading to overpopulation in high density 'hotspots' (Apollonio et al. 2010, Putman et al. 2011), resulting in these species being central within human–wildlife coexistence research. Human–ungulate coexistence is challenged across a wide variety of land uses, with severe consequences to both the species and the human population such as the damage to commercial forestry plantations (Chadwick et al. 1996, Spake et al. 2020) and crops (Linnell et al. 2020); the transmission of zoonotic diseases to livestock and eventually humans (Gortázar et al. 2012); and collisions with vehicles (Langbein et al. 2011). Most management plans depend on regulating the populations through hunting quotas, which requires a robust assessment of population densities, locally and globally (Putman et al. 2011, Krausman and Bleich 2013, Richardson et al. 2020). However, despite the importance of having accurate estimates of population densities and distributions to inform management, survey methods are rarely coordinated or standardised, and most information comes from private stakeholders' efforts to survey local populations (Apollonio et al. 2010, Liu and Nieuwenhuis 2014) or, at most, population estimates based on hunting returns (Apollonio et al. 2010, Nagy-Reis et al. 2021).

Ireland provides a representative study case to apply recent advances with ISDMs to inform ungulate management, being home to expanding populations of native red deer *Cervus elaphus*, and non-native fallow *Dama dama* and sika deer *C. nippon* (Carden et al. 2011). Despite the recent population expansion of the three species (Purser et al. 2010, Liu and Nieuwenhuis 2018), Ireland lacks a national management plan for any of its deer species and, currently, management is limited to hunting permits that do not limit hunters on where (e.g. high-density hotspots), how many and which deer (e.g. species, age and sex classes) to hunt. This is due to the lack of an empirical basis on deer distribution and relative abundance needed to set harvest quotas, maintain healthy populations and improve human–wildlife coexistence (Millspaugh et al. 2009, Williams 2011, Nagy-Reis et al. 2021). This lack of management translates into scarce and patchy data coming from a diversity of sources, which sets the perfect scene for the use of ISDMs.

We collated all data available on deer distribution in Ireland previously collected by stakeholders at different spatiotemporal scales. We also collected original data using ad hoc web tools we created and made accessible to deer stakeholders. Our goal was to demonstrate how ISDMs can integrate structured and unstructured data to produce robust predicted distributions for each species of deer present in Ireland, fundamental to inform science-based management practices. This study aimed to demonstrate the applicability of an approach that can be adapted more broadly, and ultimately produce more accurate distributions of species that can be used for science-informed wildlife conservation and for the management of human–wildlife coexistence.

Material and methods

Study species

Three species of deer are well distributed throughout Ireland: red deer, sika deer and fallow deer. Red deer are native to Ireland (but see Carden et al. 2012), whereas fallow deer were introduced by the Anglo-Normans in the 12th century (Beglane et al. 2018) and sika deer were initially introduced for ornamental purposes in the 1860s to the Powerscourt Estate, Co. Wicklow, which is situated close to the capital city of Dublin (Powerscourt 1884).

To gather all data available on deer in Ireland and Northern Ireland (NI, UK), we contacted 1) Coillte (www.coillte.ie/), which provided the results of the systematic deer PA surveys in part of the 440 000 ha of forests they manage in Ireland, and 2) the British Deer Society (https://bds.org.uk/), which provided survey data on the PA of deer in NI. These first two datasets were the only PA data available for the entire island. We collated PO data from 1) the British Agri-Food and Biosciences Institute (www.afbini.gov.uk/) which provided geotagged data on culling returns from NI. We also downloaded all observations from 2) Ireland's National Biodiversity Database (https://biodiversityireland.ie/), a citizen science platform where users can submit deer observations, 3) iNaturalist (www.inaturalist.org/), an international platform with the same goal; 4) the platform CEDaR (www.nmni.com/CEDaR/CEDaR-Centre-for-Environmental-Data-and-Recording.aspx) which curates all data for NI obtained from citizen science platforms and other surveys; and 5) the web survey (https://smartdeer.ie/) we developed ad hoc to collect PO data from Irish deer stakeholders, and where deer stalkers, other hunters, farmers and the public were asked to submit their observations (full details in Supporting information).

We obtained a total of 29 140 PA observations and 4185 PO observations, spanning between 2007 and 2022 (the vast majority being collected in the last decade, see Table 1 for full details on the temporal resolution of data). From these we generated three separate deer datasets, one for each species (red, sika and fallow deer), to run one model for each.

Data collection and pre-processing

Presence–absence (PA) data
PA data for each species were obtained from Coillte based on surveys performed in a fraction of the 6000 properties they manage (Table 1), by asking property managers (who visit the forests they manage on a regular basis) whether deer were present and, if so, what species. Properties range in size from less than 1 ha to around 2900 ha and, to assign the PA value to a specific location, we calculated the centroid of each property using the function *st_centroid()* from the package 'sf' in R (Pebesma 2018). The survey was mainly performed in 2010 and 2013, in addition to further data collected between 2014 and 2016. Some properties were surveyed only once in the period 2010–2016, but for those that were surveyed more than once, the value for that location was considered

Table 1. Summary of the presence–absence (PA, structured data) and presence-only (PO, unstructured data) datasets gathered for Ireland and Northern Ireland (NI, UK). PA data were provided by Coillte in Ireland and by the British Deer Society (BDS) in NI, while PO data were collated from a wide variety of sources including citizen science data, location of culled animals and our own web tools specifically designed for deer stakeholders (https://smartdeer.ie/).

Data type	Source	Years	Red deer	Sika deer	Fallow deer
PA	BDS survey	2016	920	920	920
	Coillte density surveys	2007–2020	417	417	417
	Coillte desk surveys	2010–2016	4936	4936	4936
	Total		6273	6273	6273
PO	Citizen science	2005–2021	408	573	394
	AFBI culling returns	2017–2021	7	169	259
	Smartdeer web survey	2021–2022	507	460	528
	Others	2001–2018	51	35	69
	Total		973	1237	1250

'absence' if deer had never been detected in the property in any of the surveys, and 'presence' in all other cases. In addition to these surveys, Coillte commissioned density surveys based on faecal pellet sampling in a subset of their properties between the years 2007 and 2020. Any non-zero densities in these data were considered 'presences', and all zeros were considered 'absences'. These data were also summarised across years when a property had been repeatedly sampled, and counted as presence if deer had been detected in any of the sampling years (Table 1).

PA data for NI were obtained from a survey carried out by the British Deer Society in 2016. The survey divided the British territory in 100 km² grid cells and deer presence or absence was assessed based on public contributions, which were then reviewed and collated by BDS experts. Since 100 km² grid cells are quite large we did not, as with the Coillte properties, calculate the centroid of each cell and assign the PA value of the cell to it. Instead, we randomly simulated positions within each cell and assigned the presence or absence value of the cell to each of them. We performed a sensitivity analysis to calculate an optimal number of positions that would capture the environmental variability within each cell (Supporting information), which was set to five random positions per grid cell. After processing, we obtained a total of 920 PA data across NI (Fig. 1A).

Presence-only (PO) data

PO data were collected from various sources, mainly (but not only) from citizen science initiatives. The National Biodiversity Data Centre (NBDC) is an Irish initiative that collates biodiversity data coming from different sources, from published studies to citizen contributions. From this repository, we obtained all contributions on the three species, a total of 1430 records. To this, we added the 164 records of deer in Ireland downloaded from the iNaturalist site, another citizen-contributed database that collects the same type of

data. From the resulting dataset, we 1) removed all observations with a spatial resolution lower than 1 km²; 2) did a visual inspection of the data and comments, and removed all observations that were obviously incorrect (i.e. at sea or that the comment specified it was a different species); 3) filtered out all the fallow deer reported in Dublin's enclosed city park (Phoenix Park) since the population there was introduced and is artificially maintained, and disconnected from the rest of populations in Ireland; and 4) filtered duplicate observations by retaining only one observation per user, location and day. The Centre for Environmental Data and Recording (CEDaR) is a data repository for NI that operates in the same way as the NBDC. They provided 872 records of deer in NI, coming from different survey, scientific and citizen science initiatives, from which we removed all records provided with a spatial resolution lower than 1 km². The location and species of 469 deer culled between 2019 and 2021 in NI were obtained from the British Agri-Food and Biosciences Institute. Some observations did not have specific coordinates, and we derived these from the location name or postcode if provided.

As part of a nationally funded initiative to improve deer monitoring in Ireland (SMARTDEER), we developed a bespoke online tool to facilitate the reporting of deer observations by the general public and all relevant stakeholders e.g. hunters, farmers or foresters. Observations were reported in 2021 and 2022 by clicking on a map to indicate a 1 km² area where deer had been observed. For each user and session, we calculated the area of the surface covered in squares, simulated a number of positions proportional to the size of the polygon and distributed them within it to generate a number of exact positions equivalent to the area where the user had indicated an observation (Supporting information). In total, the SMARTDEER tool allowed us to collect 4078 presences across Ireland and NI (Table 1, Fig. 1B).

The data used in the models were collected between 2007 and 2022. Deer populations expanded in Ireland until 2008 (Carden et al. 2011) and, according to culling return data, have somewhat stabilised since then (NPWS official data). Although the range expansion of deer species would merit further investigation, here we provide for the first time an accurate modelled distribution of the three main species of deer in Ireland and, since the data are scarce, we have made use of all available data without considering the temporal trends. A continued data collection scheme will provide enough data to study population size and range changes, but this is beyond the scope of this manuscript.

Statistical model

To integrate all the datasets (PA collected in Ireland, PA collected in NI and PO collected across the whole island) into one model for each species, we used functions from the 'PointedSDMs' package (https://github.com/PhilipMostert/PointedSDMs) to construct a joint-likelihood model. In a joint-likelihood model, a likelihood is constructed for each dataset, with shared and non-shared elements such as covariates
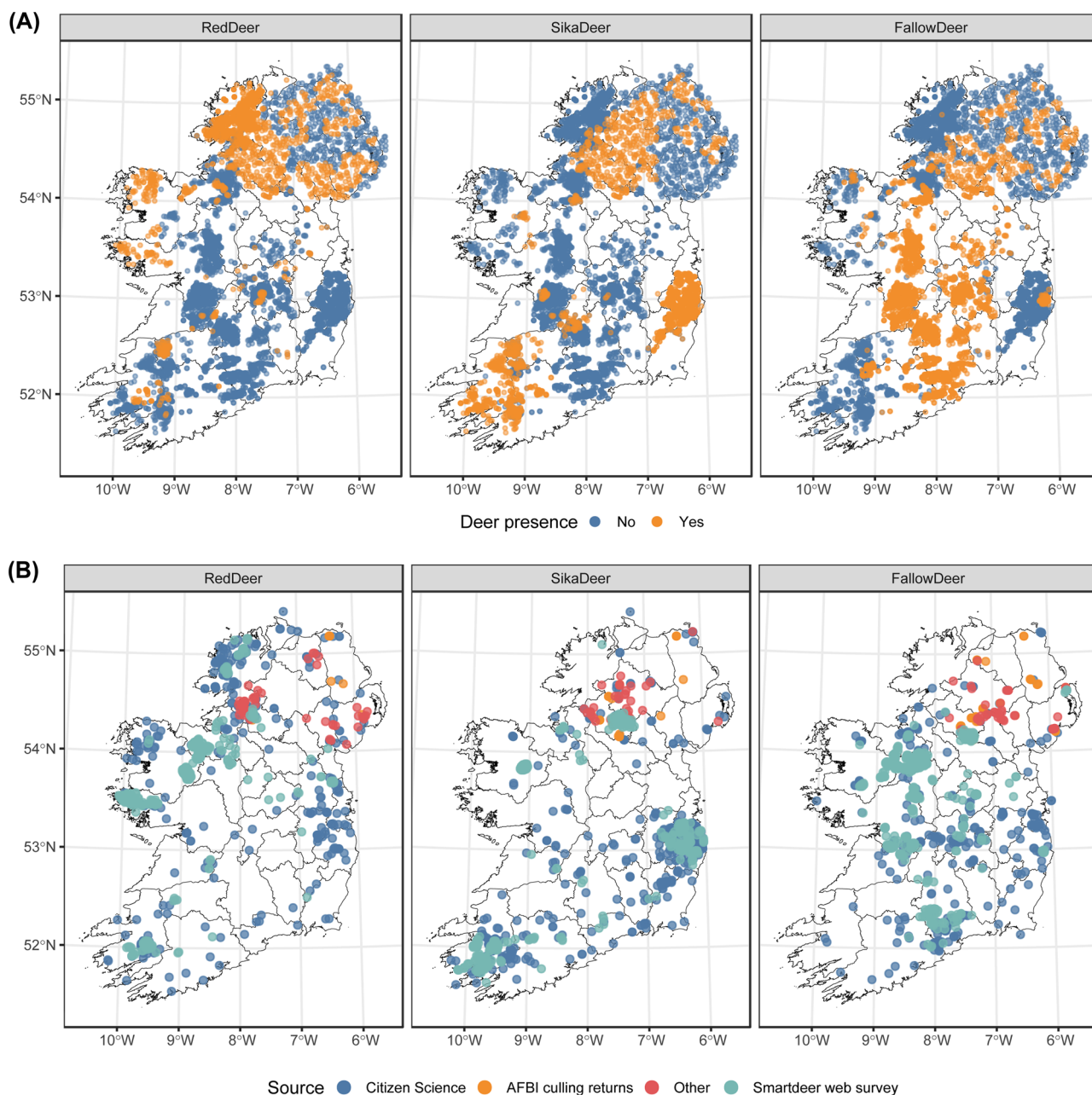
Figure 1. Presence–absence (PA, panel A) and presence-only (PO, panel B) data for each deer species (see Table 1 for sample sizes and temporal resolution). PA data were provided by Coillte in Ireland and by the British Deer Society in Northern Ireland (NI, UK), while PO data were collated from a wide variety of sources including citizen science data, location of culled animals and our own web tools specifically designed for deer stakeholders (https://smartdeer.ie/).

and spatial random effects in each of them. The likelihoods are then jointly modelled to produce a single response, a fitted object from which predictions can be made and parameters estimated. More details about the internal structure of joint models within an INLA context can be found in Blangiardo and Cameletti (2015) or Gómez-Rubio (2020).

Through the functions provided in the package, we constructed a joint model with three likelihoods: one for each of the two PA datasets (NI and Ireland, respectively), and one for the PO dataset. We considered two different likelihoods

for the PA data because of the different sampling approaches and biases associated to the two different data collected in Ireland and NI, respectively. However, we considered one single likelihood for the PO data, which were mainly generated from citizen science and thus with consistent bias across the NI–Ireland border. The three likelihoods shared all the environmental covariates (described in Covariate selection), and we included a spatial random effect that was common to the three likelihoods but with a scaling parameter (δ) between likelihoods, using the 'copy' feature of the INLA models.

The PO data were modelled as a log-Gaussian Cox process with intensity function:

$$N(A) \sim \text{Poisson} \int_A \lambda(s)\, d(s)$$

$$\log(\lambda(s)) = \alpha_1 + \beta_1 x_1 + \ldots + \beta_n x_n + \xi_s$$

where N is the expected number of presences in the study area (A), and λ is the intensity function. The linear predictor is composed of a dataset-specific intercept (α_1), a set of covariates (x_1 to x_n) with their coefficients β_1 to β_n, and a random spatial effect (spatial field) to account for the structure of the data (ξ_s). Here and in the PA likelihood we have avoided mentioning a random error term in the formula, since it was not included in the model (for simplicity; Soriano-Redondo et al. 2019, Lezama-Ochoa et al. 2020 or Laxton et al. 2022), although formally it should be included. The spatial field is modelled as Gaussian random fields with Matérn covariance functions, which is approximated using a triangulation of the study area (called a mesh) through stochastic partial differential equations fitted in a Bayesian context through integrated nested Laplace approximations (Lindgren et al. 2011).

The PA datasets were modelled as Bernouilli distributions (Isaac et al. 2020), modelling the probability of observing an individual at each location s:

Northern Ireland PA data

$$Y_s \sim \text{Bernouilli}(p_s)$$

$$\text{cloglog}(p_s) = \alpha_2 + \beta_1 x_1 + \ldots + \beta_n x_n + \delta_1 \xi_s$$

Ireland PA data

$$Z_s \sim \text{Bernouilli}(p_s)$$

$$\text{cloglog}(p_s) = \alpha_3 + \beta_1 x_1 + \ldots + \beta_n x_n + \delta_2 \xi_s$$

where Y_s and Z_s are the binary response variables (PA) and p_s is the probability of presence. These are linked to the linear predictor by a complementary log–log link function (cloglog, Ahmad Suhaimi et al. 2021). Each likelihood has its intercept (α_2 and α_3), a set of covariates (x_1 to x_n) and their coefficients (β_1 to β_n) which are common among the three likelihoods. The spatial field (ξ_s) is also common among the three likelihoods but is scaled between the PO likelihood and the two PA likelihoods by two scaling parameters (δ_1 and δ_2). Although the spatial field is scaled, its parameter estimation is common for the three likelihoods and computed from all the observations put together. Thus, the spatial field will have a common shape in the three likelihoods, but the scaling will allow for different variances in the values of the field among

likelihoods. The use of the cloglog link in the binomial model allows its response to be interpreted on the same scale as the response of the Poisson model, which allows the sharing of parameters between likelihoods (Kéry and Royle, 2016).

Prior specification

The spatial field is controlled by two hyperparameters: range and marginal variance. The range controls the smoothness of the spatial field (i.e. the distance between peaks and troughs), and the variance controls the magnitude of these peaks and troughs. In the Bayesian context in which we are fitting this model, we need to set prior values to these two hyperparameters. To do so, we use penalised complexity (PC) priors, a newly developed framework that allows easily interpretable and controllable priors (Simpson et al. 2014). PC priors are weakly informative (allowing the posterior of each hyperparameter to be mainly controlled by the data) and penalise model complexity by 'pulling' the model towards its simplest realisation (the 'base' model), which has infinite range and zero variance (i.e. a completely flat spatial field, absence of spatial structure). To set the priors, we inform the model of 'how far it is allowed to deviate' from those base models using the following specifications:

- The prior on the range (ρ) is set providing the lower tail quantile ρ_0 and the probability $P(\rho)$ so that:

$$\text{Prob}(\rho < \rho_0) = P(\rho)$$

or 'the probability that the true range (ρ) is smaller than ρ_0 is $P(\rho)$'. For example, if we set ρ_0 to be 50 and $P(\rho)$ to be 0.05, we are telling the model that the probability of the true range of the spatial field being smaller than 50 km is 5%. In this way we are limiting the range to values between infinite (the base model) and 50; that is, we are saying that the smallest that the range could possibly be is 50 km with a probability of 95%.

- The prior on the variance is set on the standard deviation (SD), providing the upper tail quantile σ_0 and the probability $P(\sigma)$ so that:

$$\text{Prob}(\sigma > \sigma_0) = P(\sigma)$$

or 'the probability that the true SD σ is larger than σ_0 is $P(\sigma)$'. For example, if we set σ_0 to be 0.5 and $P(\sigma)$ to be 0.05, the probability of the true SD being larger than 0.5 is 5%, so effectively the SD value is limited between 0 (the base model) and 0.5 with a 95% probability.

Priors have to be carefully specified, but there is no absolute rule for this, so the decisions that go into the prior choice are an essential part of the modelling process. In this case, we started off from the same point for the three species and re-ran the model with updated priors iteratively (updating the ρ_0 and σ_0 values to approximate them to the posterior mean for the range and SD, respectively) until we found a satisfactory

output (which usually meant less skewed posterior distributions for the range and SD, and for the covariate coefficients). To make the starting point as uninformative as possible (and let the data drive the posterior distributions) we set the prior of the range to 200 km with a probability of 0.5 (it is as likely that the range is above 200 km as under 200 km), and the prior for the SD as 2, with a probability of 0.5 as well (SD could either be above or under 2). In successive iterations we modified the prior values and made them more informative by reducing the probability values, until we obtained a satisfactory model. Finally, the spatial field was scaled between likelihoods by means of a scaling parameter (δ) which is not fixed beforehand, but estimated by the model, and to which we set the default priors (Gaussian distribution with mean 1 and precision 10).

Covariate selection

Raster environmental covariates used in the models were obtained from the Copernicus Land Monitoring Service (© European Union, Copernicus Land Monitoring Service 2018, European Environment Agency EEA), whereas the vector layers (roads, paths) were obtained from the Open Street Map service (OpenStreetMap contributors 2017). Planet dump (data file from January 2022, https://planet.openstreetmap.org). Vector layers were transformed into distance layers (distance to roads, distance to paths) using the *distance()* function from the package 'raster', and into density layers (density of roads, paths) using the *rasterize()* function of the same package (Hijmans 2021). All raster layers were resampled to the lowest resolution available in the used covariates, resulting in a 1 km² resolution. A full description of the process of covariate selection (including screening for collinearity) can be found in the Supporting information. The covariates eventually used in the model were elevation (m), slope (degrees), tree cover (%), small woody feature density (%), distances to forest edge (m, positive distances indicate a location outside a forest, negative distances indicate a location within a forest) and human footprint index (Venter et al. 2016, 2018; Supporting information). All covariates were scaled by subtracting the mean and dividing by the standard deviation before entering the model (function *scale()* from the *raster* package).

Spatial predictions

From the fitted models we used the *predict()* function to obtain predictions from the whole joint model in the linear scale, in a 25 km² grid. The method used to fit the models produces predictions at each location that are not a point value but a distribution, from which we can produce a posterior mean and standard deviation, thus obtaining a spatial estimate of the uncertainty of the prediction. We used the same function to obtain the prediction of the spatial effects, which can provide an indication of the spatial autocorrelation structure of each of the datasets. The model is designed on the assumption that not all individuals have been observed and, although in theory the total abundance can be calculated by predicting in the response scale and

integrating the intensity of the process over all the study area, an imperfect detection will affect the predicted total abundance. In all our models the total predicted abundances were grossly underestimated, so we decided to use the predictions in the linear scale and, rescaled from 0 to 1, used them as relative abundances instead of total abundances or densities.

Checking model performance

To check the performance of our models, we used two different datasets. On the one hand, a completely independent dataset commonly used in population estimates (Milner et al. 2006, Forsyth et al. 2022), the culling return data from the National Park & Wildlife Service (NPWS, www.npws.ie/) – responsible for issuing hunting licences – aggregated by county between 2008 and 2018. The data consist of the number of harvested deer of each species by county (ranging from 826 to 7500 km²) and year. To increase accuracy, we disregarded data before 2010 and we aggregated the data of the resulting dataset by calculating the mean number of deer harvested for each species and county, therefore during the same period of time in which we gathered the PO and PA data used in the ISDMs. Although culling data are completely independent of the datasets used in the model, they come with biases of their own. Culling returns are influenced by the number of licences issued by each country and, ultimately, by their human population size. In addition, in some counties where national parks are present, the culling returns are lower than would be expected based on deer abundance data, since hunting is not allowed within national parks. Lastly, spatial distribution of culling returns may be affected by hunters' customs and traditions, which might cause them to prefer hunting in some counties rather than others. Despite these biases, culling returns are usually considered a valuable snapshot of deer populations (Milner et al. 2006, Nagy-Reis et al. 2021), and we decided to challenge our ISDMs model predictions by correlating them to independent culling returns.

On the other hand, we used the density estimates obtained through the faecal pellet counts within Coillte properties (Presence–absence (PA) data). This dataset is less biased compared to culling returns and is well spread within Coillte properties across Ireland, although it is not entirely independent to a small subset (8.5%) of the PA data used in the ISDMs, which were converted to PA to estimate PA likelihood of the model. Whenever a property was sampled more than once, the average density among all sampling instances was used.

With the predictions obtained from the ISDM model we 1) aggregated the value of the predictions by county (to be correlated to county-level culling returns), and 2) extracted the value of the predictions at the centroid of each Coillte property (to be correlated to property-level faecal pellet densities). We tested the correlation of our predictions to these two datasets with a non-parametric Kendall's tau (Hollander and Wolfe 1973), selected based on lack of normality and non-linearity in the data series.

Results

We developed one model for each species, including effects for six covariates (tree cover, density of small woody features, distance to the forest edge, slope, elevation and human footprint index), and a shared spatial field for each species model, scaled between the likelihoods (Supporting information). When run with the initial uninformative priors, the models produced estimates for the covariate effect sizes, and the range and standard deviation of the spatial field, that were skewed and with large uncertainties. Through iteratively running the model with increasingly informative priors we obtained fitted models for each species that produced centred and more certain estimates of all parameters (Table 2).

The covariate effects for the three models (Fig. 2) showed that the three species had, in general, similar ecology in terms of environmental preferences. Red, sika and fallow deer were more likely to be observed within forests (negative values of distance to forest edge) with high tree cover densities. Elevation had no significant effect in the distribution of red and fallow deer (95% credible intervals, CI, overlapping 0), and a very small positive effect on the distribution of sika deer, while slope did not have a significant effect in red deer but opposite effects in sika and fallow, with the former preferring steeper and the latter flatter terrain. The three species distributions seemed to match areas with greater human footprint, although the effect was only significant in red deer, in line with the expectation

that bare and unpopulated lands are less attractive to deer. The spatial fields for the three species showed posterior ranges between 94 and 228 km, and posterior SD between 0.49 and 2.32 (Supporting information, Table 2).

From each of the models we obtained a spatial prediction that allowed us to plot its mean and its standard deviation (Fig. 3). Red deer hotspots were detected in the north-west and south-west of Ireland, with a smaller, more diffuse hotspot to the north-east. Sika deer were present at higher relative abundances in a hotspot at the east coast and in the south-west, with again less intense, more diffuse populations in the north-west, frequently overlapping with the red deer. Last, fallow deer are mainly distributed in the midlands, with high abundances also in the north of Ireland and in NI. Standard deviation was comparatively larger for red deer than for the other two species, likely related to the fact that this is the less abundant and more sparsely distributed species (Fig. 3).

Our ISDMs predicted the distribution of the three species across Ireland and, when aggregated by county and compared to the county-level culling returns, showed a relatively high correlation (Kendall correlation values between 0.45 and 0.60, Table 3, Supporting information) with the independent dataset of culling returns corrected by hunters' licences. When compared to the densities calculated from faecal pellets within Coillte properties, the agreement was also relatively high (Kendall correlation values between 0.45 and 0.64, Table 3, Supporting information).

Table 2. Priors' specification and posterior distribution estimated for the spatial fields and the scaling parameters of the three species-specific models (red, sika and fallow deer). Spatial fields are defined in our model by their range (ρ) and standard deviation (σ), and the scaling parameters between likelihoods (δ_1 and δ_2). Priors are set on the range and standard deviation as penalised complexity (PC) priors (Materials and methods) and so the lower limit for the range and the upper limit for the SD are specified, together with a probability value for each. For the scaling parameters, priors were left by default as a Gaussian distribution with mean 1 and precision 10. The posterior distributions of these parameters were obtained from the model, and are described here by their mean and standard deviation. PA, presence–absence; PO, presence-only; NI, Northern Ireland.

| Species	Parameter	Prior	Posterior mean	Posterior SD
Red deer	Range of the spatial field	$\rho_0 = 100, P(\rho) = 0.1$	97.17	0.357
	Standard deviation of the spatial field	$\sigma_0 = 0.5, P(\sigma) = 0.1$	0.49	0.001
	Scaling parameter (PO to PA-NI)	$\lambda_1 \sim N(1, \frac{1}{\sqrt{10}})$	1.08	0.003
	Scaling parameter (PO to PA-Ireland)	$\lambda_2 \sim N(1, \frac{1}{\sqrt{10}})$	2.14	0.003
Sika deer	Range of the spatial field	$\rho_0 = 200, P(\rho) = 0.1$	228.54	25.488
	Standard deviation of the spatial field	$\sigma_0 = 0.1, P(\sigma) = 0.1$	1.04	0.058
	Scaling parameter (PO to PA-NI)	$\lambda_1 \sim N(1, \frac{1}{\sqrt{10}})$	0.20	0.026
	Scaling parameter (PO to PA-Ireland)	$\lambda_2 \sim N(1, \frac{1}{\sqrt{10}})$	0.43	0.017
Fallow deer	Range of the spatial field	$\rho_0 = 100, P(\rho) = 0.5$	148.72	0.203
	Standard deviation of the spatial field	$\sigma_0 = 2, P(\sigma) = 0.1$	2.32	0.004
	Scaling parameter (PO to PA-NI)	$\lambda_1 \sim N(1, \frac{1}{\sqrt{10}})$	−0.17	0.001
	Scaling parameter (PO to PA-Ireland)	$\lambda_2 \sim N(1, \frac{1}{\sqrt{10}})$	0.84	0.003
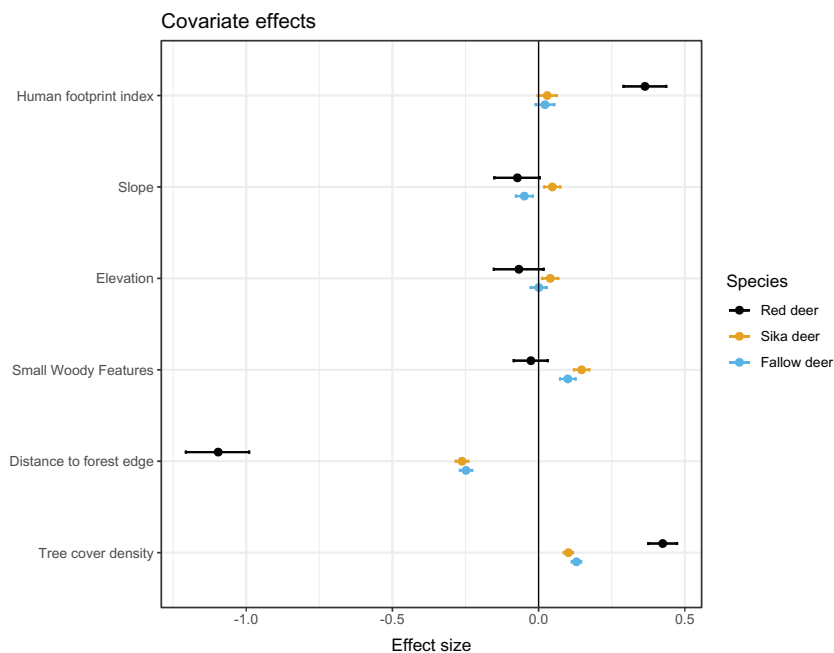
Covariate effects

Figure 2. Covariate effects for each of the models for red (top, black), sika (middle, orange) and fallow (bottom, blue) deer. Circles represent the median value of the effect, while the bars represent the 95% credible intervals (Cis).

Discussion

Applicability of joint likelihood models in an INLA context to real data

Our results demonstrated the practical applications of ISDM in the INLA Bayesian context despite the scarcity and low quality of the data. Our models managed to successfully produce not only a predicted distribution for each species, but also to map their uncertainty and estimate relative abundance. The predicted distributions displayed low uncertainties (reflected in small standard deviations), but it was higher for the red deer model, efficiently reflecting the most data-deficient species (i.e. fewer data relate to less certain predictions). Furthermore, we checked the performance of the predictions with two different datasets to ensure their accuracy, finding that our models performed well when compared to both county-level culling returns and deer densities estimated from faecal pellet counts. Thus, we provide accurate science-based distribution maps that integrate all previous knowledge about deer distribution in Ireland, setting a path for future data gathering initiatives with conservation and management in sight.

Deer distributions and relative abundances in Ireland

Our model predicted several population centres for red deer. One of the population predicted hotspots was centred around the Killarney National Park, a herd under conservation measures such as a hunting ban in the area (Carden et al. 2012). The other hotspots to the north-west coincided with areas

where modern introductions of red deer have taken place in the past two centuries (Purser et al. 2010), and the diffuse populations along the east coast correspond to the area where the first recorded introduction of red deer into Ireland took place in 1246 (McDevitt et al. 2009).

The sika deer model showed two very clear hotspots in the east and south-west of the island, and two less dense populations in the north-west, reflecting the history of their introduction in Ireland (Purser et al. 2010). There was considerable overlap between the populations of red and sika deer, which could merit further study on their habitat and diet preferences to investigate the possible niche, spatial or temporal segregation that might facilitate coexistence. From our covariate effects, sika and red deer seemed to have similar preferences for tree-dense areas within the forests, although the distribution of red deer seemed to be much more driven by the environment than that of sika deer, which could be explained by the fact that sika deer were artificially introduced, and thus their distribution may be driven more by a founder effect (they are more abundant in or near the places they were introduced) than by their environmental preferences (Sherpa et al. 2020). Although sika deer did seem to have a preference for higher, steeper areas and for small woody features, whichred deer do not display (Cis overlapping 0), the distribution overlap of the two species causes concerns with regards to the hybridisation between the two. This has been observed both in captivity and in the wild (Abernethy 1994) and could be a threat to the genetic purity of the Kerry herd (Smith et al. 2014).

Fallow deer were predicted to be distributed mostly over the areas from where the other two species were largely absent. This might be due to different habitat and food preferences,
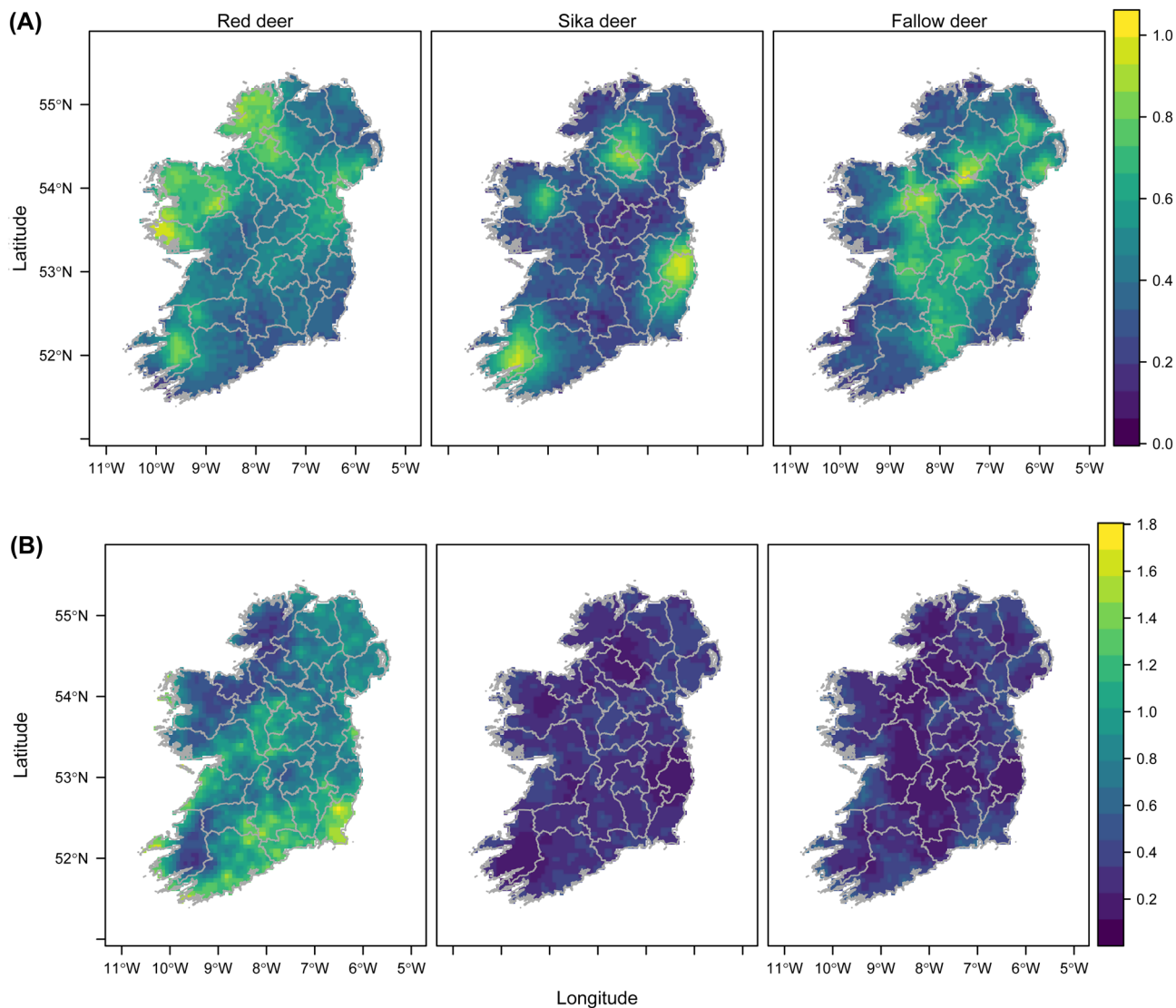
Figure 3. Mean (panel (A)) and standard deviation (panel (B)) of the spatial predictions for red, sika and fallow deer. The values indicate relative abundances, with 0 reflecting absence of the species and values closer to 1 representing the areas where the species is more abundant.

since fallow deer are known to be more obligated grazers than either red or sika deer (Obidziński et al. 2013), or to competitive exclusion, but could also be a reflection of the founder effect since fallow deer seem to have slow range expansion rates from where their populations are first established (Ward 2005). Nevertheless, since the last published distribution in 2008 (Carden et al. 2011), fallow deer distribution seems to have expanded northward, now displaying a continuous distribution from the south-east coast through the midlands and the west and all the way up to the north-west coast and NI.

For each model, the spatial field shared among the three likelihoods captured the spatial structure not explained by the covariates, as well as any spatial autocorrelation among the sampled locations, with higher values where data were more clustered (Adde et al. 2021). For the three models, the spatial field contributed greatly to the overall predicted distribution, with the covariates having in general small effects for all species. In other words, the fact that the spatial field (and not the environmental covariates) captured most of the spatial structure in the data suggests that environmental variables are not the main drivers of the distributions reflecting the 'founder effect' in their distribution, since the

Table 3. Kendall correlation coefficients between the relative abundances of red, sika and fallow deer as predicted by our integrated species distribution models (ISDMs) and 1) culling returns data at the Irish county level (n=26) and 2) deer densities from faecal pellet counts at the Coillte property level (n=the prediction values extracted at the Coillte property centroids and the deer densities estimated in them by faecal pellet counts (n=417).

	Culling returns	Density from pellets
Red deer	0.60	0.45
Sika deer	0.45	0.64
Fallow deer	0.46	0.59

three species have been introduced (or re-introduced in the case of the red deer; Powerscourt 1884, Carden et al. 2012, Beglane et al. 2018).

Joint likelihood models as a tool for management in data-scarce scenarios

Our predicted distributions described an island where deer of at least one species were omnipresent, with some regions where two species spatially overlap. The covariates showed that although the three species preferred areas with dense tree cover and within forests, that did not necessarily mean that deer shy away from human presence, reflected in our models by a positive effect of human footprint index, particularly for red deer. That is, however, more reflective of Ireland's natural habitats than of deer preferences: Ireland and NI have a large proportion of heavily modified habitat (approximately 69% of Ireland and 76% of NI are covered by farmland) (Central Statistics Office 2021, Northern Ireland Statistics and Research Agency 2021), with most of their agricultural land devoted to permanent and rough grazing grasslands, very attractive to deer (Drennan et al. 2005, O'Mara 2012), The forests, small and patchily distributed, are mostly non-native and are present within mosaics dominated by human-modified habitats, making it almost impossible for deer to avoid anthropomorphised environments. This has obvious consequences for human–wildlife coexistence, since deer have more opportunities to interact heavily with human resources such as roads, commercial forestry and farms. Thus, these results constitute a starting point for management, by providing information on areas where the relative densities of the relevant deer species are higher, and where targeted actions would be most effective.

With this research, we have demonstrated the applicability of joint Bayesian spatial models fitted through INLA methods to obtain accurate distributions and relative abundances of species, even when data are scarce and with diverse biases. Our models' performance has been checked with two different datasets, proving their accuracy even with low-quality, patchy data, which makes them a useful tool for the management and conservation of wildlife in most contexts where a data collection protocol has not been established. Our work now opens new and exciting future scenarios, because the same type of model can be adapted to estimate actual abundances by including data on the number of individuals (e.g. group sizes) and sampling effort, leading ISDMs to produce even more accurate information on species abundances which are so essential for science-informed management.

Author contributions

Virginia Morera-Pujol: Conceptualization (equal); Data curation (equal); Formal analysis (lead); Methodology (lead); Validation (lead); Visualization (lead); Writing – original draft (lead); Writing – review and editing (lead). **Philip S. Mostert**: Formal analysis (supporting); Methodology (supporting); Software (lead); Writing – review and editing (supporting). **Kilian J. Murphy**: Conceptualization (equal); Data curation (equal); Writing – review and editing (supporting). **Tim Burkitt**: Data curation (equal); Writing – review and editing (supporting). **Barry Coad**: Data curation (equal); Writing – review and editing (supporting). **Barry J. McMahon**: Conceptualization (supporting); Project administration (supporting); Writing – review and editing (supporting). **Maarten Nieuwenhuis**: Conceptualization (supporting); Project administration (supporting); Writing – review and editing (supporting). **Kevin Morelle**: Software (equal); Writing – review and editing (supporting). **Alastair I. Ward**: Data curation (equal); Writing – review and editing (supporting). **Simone Ciuti**: Conceptualization (equal); Data curation (equal); Funding acquisition (equal); Project administration (equal); Supervision (equal); Writing – review and editing (supporting).

Transparent peer review

The peer review history for this article is available at https://publons.com/publon/10.1111/ecog.06451.

Data availability statement

Data are available from the Dryad Digital Repository: https://10.5061/dryad.5mkkwh795 (Morera-Pujol et al. 2022).

Supporting information

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

References

Abernethy, K. 1994. The establishment of a hybrid zone between red and sika deer (genus *Cervus*). – Mol. Ecol. 3: 551–562.

Adde, A., Casabona i Amat, C., Mazerolle, M. J., Darveau, M., Cumming, S. G. and O'Hara, R. B. 2021. Integrated modeling of waterfowl distribution in western Canada using aerial survey and citizen science (eBird) data. – Ecosphere 12: e03790.

Ahmad Suhaimi, S. S., Blair, G. S., Jarvis, S. G. and Leroy, B. 2021. Integrated species distribution models: a comparison of approaches under different data quality scenarios. – Divers. Distrib. 27: 1066–1075.

Apollonio, M., Andersen, R. and Putman, R. J. 2010. European ungulates and their management in the 21st century. – Cambridge Univ. Press.

Bachl, F. E., Lindgren, F., Borchers, D. L. and Illian, J. B. 2019. inlabru: an R package for Bayesian spatial modelling from ecological survey data. – Methods Ecol. Evol. 10: 760–766.

Bakka, H., Rue, H., Fuglstad, G. A., Riebler, A., Bolin, D., Illian, J. B., Krainski, E., Simpson, D. P. and Lindgren, F. K. 2018. Spatial modeling with R-INLA: a review. – Wiley Interdiscip. Rev. Comput. Stat. 10: 1–24.

Beglane, F., Baker, K., Carden, R. F., Hoelzel, A. R., Lamb, A. L., Fhionnghaile, R. M., Miller, H. and Sykes, N. 2018. Ireland's fallow deer: their historical, archaeological and biomolecular records. – Proc. R. Irish Acad. Archaeol. Cult. Hist. Literat. 118C: 141–165.

Blangiardo, M. and Cameletti, M. 2015. Spatial and spatio-temporal Bayesian models with R-INLA. – Wiley.

Boyce, M. S. and Corrigan, R. 2017. Moose survey app for population monitoring. – Wildl. Soc. Bull. 41: 125–128.

Carden, R. F., Carlin, C. M., Marnell, F., McElholm, D., Hetherington, J. and Gammell, M. P. 2011. Distribution and range expansion of deer in Ireland. – Mammal Rev. 41: 313–325.

Carden, R. F., McDevitt, A. D., Zachos, F. E., Woodman, P. C., O'Toole, P., Rose, H., Monaghan, N. T., Campana, M. G., Bradley, D. G. and Edwards, C. J. 2012. Phylogeographic, ancient DNA, fossil and morphometric analyses reveal ancient and modern introductions of a large mammal: the complex case of red deer *Cervus elaphus* in Ireland. – Quater. Sci. Rev. 42: 74–84.

Carlson, C. J. 2020. embarcadero: species distribution modelling with Bayesian additive regression trees in R. – Methods Ecol. Evol. 11: 850–858.

Central Statistics Office 2021. Environmental indicators Ireland 2021. – www.cso.ie.

Chadwick, A. H., Ratcliffe, P. R. and Abernethy, K. 1996. Sika deer in Scotland: density, population size, habitat use and fertility – some comparisons with red deer. – Scottish Forestry 50: 8–16.

Cunningham, C. X., Comte, S., McCallum, H., Hamilton, D. G., Hamede, R., Storfer, A., Hollings, T., Ruiz-Aravena, M., Kerlin, D. H., Brook, B. W., Hocking, G. and Jones, M. E. 2021. Quantifying 25 years of disease-caused declines in Tasmanian devil populations: host density drives spatial pathogen spread. – Ecol. Lett. 24: 958–969.

Dorazio, R. M. 2014. Accounting for imperfect detection and survey bias in statistical analysis of presence-only data. – Global Ecol. Biogeogr. 23: 1472–1484.

Drennan, M. J., Carson, A. F. and Crosse, S. 2005. Overview of animal production from pastures in Ireland. – In: Murphy, J. J. (ed.), Utilisation of grazed grass in temperate animal systems. Proceedings of a satellite workshop of the XXth International Grassland Congress, July 2005, Cork, Ireland. Wageningen Academic Publishers, pp. 19–35.

Elith, J. and Leathwick, J. R. 2009. Species distribution models: ecological explanation and prediction across space and time. – Annu. Rev. Ecol. Evol. Syst. 40: 677–697.

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E. and Yates, C. J. 2011. A statistical explanation of MaxEnt for ecologists. – Divers. Distrib. 17: 43–57.

Fletcher Jr, R. J., Hefley, T. J., Robertson, E. P., Zuckerberg, B., McCleery, R. A. and Dorazio, R. M. 2019. A practical guide for combining data to model species distributions. – Ecology 100: e02710.

Forsyth, D. M., Comte, S., Davis, N. E., Bengsen, A. J., Côté, S. D., Hewitt, D. G., Morellet, N. and Mysterud, A. 2022. Methodology matters when estimating deer abundance: a global systematic review and recommendations for improvements. – J. Wildl. Manage. 86: e22207.

Frans, V. F., Augé, A. A., Fyfe, J., Zhang, Y., McNally, N., Edelhoff, H., Balkenhol, N. and Engler, J. O. 2021. Integrated SDM database: enhancing the relevance and utility of species distribution models in conservation management. – Methods Ecol. Evol. 13: 243–261.

Gómez-Rubio, V. 2020. Bayesian inference with INLA. – Chapman & Hall/CRC Press.

Gortázar, C., Delahay, R. J., McDonald, R. A., Boadella, M., Wilson, G. J., Gavier-Widen, D. and Acevedo, P. 2012. The status of tuberculosis in European wild mammals. – Mammal Rev. 42: 193–206.

Guillera-Arroita, G., Lahoz-Monfort, J. J., Elith, J., Gordon, A., Kujala, H., Lentini, P. E., McCarthy, M. A., Tingley, R. and Wintle, B. A. 2015. Is my species distribution model fit for purpose? Matching data and models to applications. – Global Ecol. Biogeogr. 24: 276–292.

Guisan, A. and Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. – Ecol. Lett. 8: 993–1009.

Guisan, A. and Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. – Ecol. Model. 135: 147–186.

Hijmans, R. J. 2021. raster: geographic data analysis and modeling. – R package ver. 3.4-10, https://CRAN.R-project.org/package=raster.

Hollander, M. and Wolfe, D. A. 1973. Nonparametric statistical methods. – Wiley, pp. 185–194.

Hortal, J. and Lobo, J. M. 2005. An ED-based protocol for optimal sampling of biodiversity. – Biodivers. Conserv. 14: 2913–2947.

Isaac, N. J. B., Jarzyna, M. A., Keil, P., Dambly, L. I., Boersch-Supan, P. H., Browning, E., Freeman, S. N., Golding, N., Guillera-Arroita, G., Henrys, P. A., Jarvis, S., Lahoz-Monfort, J., Pagel, J., Pescott, O. L., Schmucki, R., Simmonds, E. G. and O'Hara, R. B. 2020. Data integration for large-scale models of species distributions. – Trends Ecol. Evol. 35: 56–67.

Kéry, M. and Royle, J. A. 2015. Applied hierarchical modeling in ecology: analysis of distribution, abundance and species richness in R and BUGS. Academic Press, https://doi.org/10.1016/C2015-0-04070-9.

Krausman, P. R. and Bleich, V. C. 2013. Conservation and management of ungulates in North America. – Int. J. Environ. Stud. 70: 372–382.

Langbein, J., Putman, R. and Pokorny, B. 2011. Traffic collisions involving deer and other ungulates in Europe and available measures for mitigation. – In: Apollonio, M., Andersen, R. and Putman, R. (eds), Ungulate management in Europe: problems and practices. Cambridge Univ. Press, pp. 215–259.

Laxton, M. R., Rodríguez de Rivera, O., Soriano-Redondo, A. and Illian, J. B. 2022. Balancing structural complexity with ecological insight in spatio-temporal species distribution models. – Methods Ecol. Evol. https://doi.org/10.1111/2041-210X.13957.

Lezama-Ochoa, N., Pennino, M. G., Hall, M. A., Lopez, J. and Murua, H. 2020. Using a Bayesian modelling approach (INLA-

SPDE) to predict the occurrence of the spinetail devil ray (*Mobular mobular*). – Sci. Rep. 10: 18822.

Lindgren, F. and Rue, H., 2015. Bayesian spatial modelling with R-INLA. – J. Stat. Softw. 63: 1–25.

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. – J. R. Stat. Soc. B 73: 423–498.

Linnell, J. D. C. and Zachos, F. E. 2011. Status and distribution patterns of European ungulates: genetics, population history and conservation. – In: Apollonio, M., Andersen, R. and Putman, R. (eds), Ungulate management in Europe: problems and practices. Cambridge Univ. Press, pp. 12–53.

Linnell, J. D. C., Cretois, B., Nilsen, E. B., Rolandsen, C. M., Solberg, E. J., Veiberg, V., Kaczensky, P., Van Moorter, B., Panzacchi, M., Rauset, G. R. and Kaltenborn, B. 2020. The challenges and opportunities of coexisting with wild ungulates in the human-dominated landscapes of Europe's Anthropocene. – Biol. Conserv. 244: 108500.

Liu, Y. and Nieuwenhuis, M. 2014. An analysis of habitat-use patterns of fallow and sika deer based on culling data from two estates in Co. Wicklow. – Irish Forestry 71: 27–49.

Liu, Y. and Nieuwenhuis, M. 2018. The effect of roadside land-use on the occurrence of deer vehicle collisions. – Irish Forestry 75: 8–25.

Lobo, J. M., Jiménez-Valverde, A. and Hortal, J. 2010. The uncertain nature of absences and their importance in species distribution modelling. – Ecography 33: 103–114.

McDevitt, A. D., Edwards, C. J., O'Toole, P., O'Sullivan, P., O'Reilly, C. and Carden, R. F. 2009. Genetic structure of, and hybridisation between, red *Cervus elaphus* and sika *Cervus nippon* deer in Ireland. – Mammal. Biol. 74: 263–273.

Miller, D. A. W., Pacifici, K., Sanderlin, J. S., Reich, B. J. and Gardner, B. 2019. The recent past and promising future for data integration methods to estimate species' distributions. – Methods Ecol. Evol. 10: 22–37.

Millspaugh, J. J., Skalski, J. R., Townsend, R. L., Diefenbach, D. R., Boyce, M. S., Hansen, L. P. and Kammermeyer, K. 2009. An evaluation of sex–age–kill (SAK) model performance. – J. Wildl. Manage. 73: 442–451.

Milner, J. M., Bonenfant, C., Mysterud, A., Gaillard, J.-M., Csányi, S. and Stenseth, N. C. 2006. Temporal and spatial development of red deer harvesting in Europe: biological and cultural factors. – J. Appl. Ecol. 43: 721–734.

Morera-Pujol, V., Mostert, P. S., Murphy, K., Burkitt, T., Coad, B., McMahon, B. J., Nieuwenhuis, M., Morelle, K., Ward, A. and Ciuti, S. 2022. Data from: Bayesian species distribution models integrate presence-only and presence–absence data to predict deer distribution and relative abundance. – Dryad Digital Repository, https://doi.org/10.1111/ecog.06451.

Nagy-Reis, M., Reimer, J. R., Lewis, M. A., Jensen, W. F. and Boyce, M. S. 2021. Aligning population models with data: adaptive management for big game harvests. – Global Ecol. Conserv. 26: e01501.

Northern Ireland Statistics and Research Agency 2021. Statistical review of Northern Ireland Agriculture 2020. – Dept of Agriculture, Environment and Rural Affairs.

Nyhus, P. J. 2016. Human–wildlife conflict and coexistence. – Annu. Rev. Environ. Resour. 41: 143–171.

O'Mara, F. P. 2012. The role of grasslands in food security and climate change. – Ann. Bot. 110: 1263–1270.

Obidziński, A., Kiełtyk, P., Borkowski, J., Bolibok, L. and Remuszko, K. 2013. Autumn–winter diet overlap of fallow, red and roe deer in forest ecosystems, Southern Poland. – Open Life Sci. 8: 8–17.

Open Street Map Contributors 2017. Planet dump. – https://planet.openstreetmap.org.

Pacifici, K., Reich, B. J., Miller, D. A. W., Gardner, B., Stauffer, G., Singh, S., McKerrow, A. and Collazo, J. A. 2017. Integrating multiple data sources in species distribution modeling: a framework for data fusion. – Ecology 98: 840–850.

Paradinas, I., Conesa, D., López-Quílez, A. and Bellido, J. M. 2017. Spatio-temporal model structures with shared components for semi-continuous species distribution modelling. – Spatial Stat. 22: 434–450.

Pebesma, E. J. 2018. Simple features for R: standardized support for spatial vector data. – R J. 10: 439–446.

Powerscourt, V. 1884. On the acclimatization of the Japanese deer at Powerscourt. – Proc. Zool. Soc. Lond. 52: 207–209.

Purser, P., Wilson, F. and Carden, R. F. 2010. Deer and forestry in Ireland: a review of current status and management requirements. – Coillearnacha Dúchasacha (Woodlands of Ireland) Steering Committee.

Putman, R., Apollonio, M. and Andersen, R. 2011. Ungulate management in Europe: problems and practices. – Cambridge Univ. Press.

Richardson, S., Mill, A. C., Davis, D., Jam, D. and Ward, A. I. 2020. A systematic review of adaptive wildlife management for the control of invasive, non-native mammals and other human–wildlife conflicts. – Mammal Rev. 50: 147–156.

Rue, H., Martino, S. and Chopin, N. 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. – J. R. Stat. Soc. B 71: 319–392.

Sherpa, S., Renaud, J., Guéguen, M., Besnard, G., Mouyon, L., Rey, D. and Després, L. 2020. Landscape does matter: disentangling founder effects from natural and human-aided post-introduction dispersal during an ongoing biological invasion. – J. Anim. Ecol. 89: 2027–2042.

Simmonds, E. G., Jarvis, S. G., Henrys, P. A., Isaac, N. J. B. and O'Hara, R. B. 2020. Is more data always better? A simulation study of benefits and limitations of integrated distribution models. – Ecography 43: 1413–1422.

Simpson, D. P., Rue, H., Martins, T. G., Riebler, A. and Sørbye, S. H. 2014. Penalising model component complexity: a principled, practical approach to constructing priors. – Stat. Sci. 32: 1–42.

Smith, S. L., Carden, R. F., Coad, B., Birkitt, T. and Pemberton, J. M. 2014. A survey of the hybridisation status of *Cervus* deer species on the island of Ireland. – Conserv. Genet. 15: 823–835.

Soriano-Redondo, A., Jones-Todd, C. M., Bearhop, S., Hilton, G. M., Lock, L., Stanbury, A., Votier, S. C. and Illian, J. B. 2019. Understanding species distribution in dynamic populations: a new approach using spatio-temporal point process models. – Ecography 42: 1092–1102.

Spake, R., Bellamy, C., Gill, R., Watts, K., Wilson, T., Ditchburn, B., Eigenbrod, F. and Bennett, J. 2020. Forest damage by deer depends on cross-scale interactions between climate, deer density and landscape structure. – J. Appl. Ecol. 57: 1376–1390.

Venter, O., Sanderson, E. W., Magrach, A., Allan, J. R., Beher, J., Jones, K. R., Possingham, H. P., Laurance, W. F., Wood, P., Fekete, B. M., Levy, M. A. and Watson, J. E. 2016. Sixteen years of change in the global terrestrial human footprint and

implications for biodiversity conservation. – Nat. Commun. 7: 12558.

Venter, O., Sanderson, E. W., Magrach, A., Allan, J. R., Beher, J., Jones, K. R., Possingham, H. P., Laurance, W. F., Wood, P., Fekete, B. M., Levy, M. A. and Watson, J. E. 2018. Last of the wild project, ver. 3 (LWP-3): 2009 human footprint, 2018 release. – NASA Socioeconomic Data and Applications Center (SEDAC).

Ward, A. I. 2005. Expanding ranges of wild and feral deer in Great Britain. – Mammal Rev. 35: 165–173.

Williams, B. K. 2011. Adaptive management of natural resources – framework and issues. – J. Environ. Manage. 92: 1346–1353.

Yackulic, C. B., Chandler, R., Zipkin, E. F., Royle, J. A., Nichols, J. D., Campbell Grant, E. H., Veran, S. and O'Hara, R. B. 2013. Presence-only modelling using MAXENT: when can we trust the inferences? – Methods Ecol. Evol. 4: 236–243.

Zulian, V., Miller, D. A. W. and Ferraz, G. 2021. Integrating citizen-science and planned-survey data improves species distribution estimates. – Divers. Distrib. 27: 2498–2509.