



EXTERNAL SCIENTIFIC REPORT

APPROVED: 3 February 2022

doi:10.2903/sp.efsa.2022.EN-7174

Update of model for wild ruminant abundance based on occurrence and first models based on hunting yield at European scale

ENETWILD-consortium¹, Sonia Illanas, Simon Croft, Pelayo Acevedo, Javier Fernández-López, Joaquín Vicente, Jose Antonio Blanco-Aguiar, Roberto Pascual-Rico, Massimo Scandura, Marco Apollonio, Ezio Ferroglio, Oliver Keuling, Stefania Zanet, Tomasz Podgorski, Kamila Plis, Francesca Brivio, Carmen Ruiz, Ramon C. Soriguer, Rachele Vada, Graham C. Smith.

Abstract

In a previous report, ENETWILD proposed a generic model framework to predict habitat suitability and likely occurrence for wild ruminant species using opportunistic presence data (occurrence records for wild ungulate species from the Global Biodiversity Information Facility). In this report, for the first time, we develop models based on hunting yield data (HY) for the most widely distributed wild ruminant species in Europe: roe deer (*Capreolus capreolus*) and red deer (*Cervus elaphus*). We also update models based on occurrence (roe deer, red deer, fallow deer (*Dama dama*), European moose (*Alces alces*) and muntjac (*Muntiacus reevesi*), evaluate the performance of both approaches, and compare outputs. As for HY models, we could not conduct one model per bioregion as there are not enough data for modelling in some bioregions, and therefore, we calibrated a unique model, including eco-geographical variables as predictors. The calibration plots for HY models showed a good predictive performance for red deer in the Eastern bioregion and roe deer at Eastern and Western. The abundance distribution pattern of red deer HY was widely scattered over all Europe, as expected for a widely distributed species which shows high ecological plasticity, and roe deer presented the highest abundance in Atlantic and Eastern Europe, progressively decreasing towards Northern Mediterranean bioregions. Overall, calibration plot did not perform well in the Northern region, which could be due to the low availability of data for both species in this bioregion. As for occurrence data models, performances using our revised approach for most species showed similarly moderate predictive accuracy. To sum, HY model projections showed good patterns where good quality data was provided, while worst predictions are found in neighbouring countries/bioregions. Two approximations to be explored for next models are: (i) modelling HY per bioregion providing more flexibility to the models, even if data projection is done at lower resolution scales, and (ii), modelling HY by accounting the fact that certain countries provide most data, to avoid that these areas overinform the model. As for occurrence data model, next steps for data acquisition and occurrence data modelling are: (i) review target group definitions for each species, (ii) revise definitions of "true" absence for model testing for better parity with fitting, and (iii) either replace principal component analysis with variance inflation factor analysis to remove co-correlates and model calibration for variable selection or develop post-model analysis to recover environmental dependencies.

© European Food Safety Authority, 2022

¹ ENETWILD Consortium: www.enetwild.com

Key words: distribution, hunting bags, occurrence, population abundance, population monitoring, risk assessment, spatial modelling, wild ruminants, red deer, roe deer, *Cervus elaphus*, *Capreolus capreolus*

Question number: EFSA-Q-2020-00679

Correspondence: biohaw@efsa.europa.eu

Disclaimer: The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Acknowledgements: We acknowledge EFSA ALPHA, and ENETWILD partners for reviewing this manuscript. We are grateful to ENETWILD collaborators (listed on <http://www.enetwild.com/collaborators>) and data providers and institutions feeding GBIF (listed at <https://doi.org/10.15468/dl.up2acp>) for sharing data on wild ungulates.

Suggested citation: ENETWILD-consortium², Illanas S, Croft S, Acevedo P, Fernández-López J, Vicente J, Blanco-Aguiar JA, Pascual-Rico R, Scandura M, Apollonio M, Ferroglio E, Keuling O, Zanet S, Podgorski T, Plis K, Brivio F, Ruiz C, Soriguer RC, Vada R, Smith GC, 2022. Update of model for wild ruminant abundance based on occurrence and first models based on hunting yield at European scale. 2022:EN-7174. 30 pp. doi:10.2903/sp.efsa.2022.EN-7174

ISSN: 2397-8325

© European Food Safety Authority, 2022

Reproduction is authorised provided the source is acknowledged.

² ENETWILD Consortium: www.enetwild.com

Summary

Background and objectives: In a previous report (ENETWILD consortium et al., 2021), ENETWILD proposed a generic model framework to predict habitat suitability and likely occurrence for wild ruminant species using opportunistic presence data (occurrence records for wild ungulate species from the Global Biodiversity Information Facility). Across all wild ungulate species (roe deer *Capreolus capreolus*, red deer *Cervus elaphus*, fallow deer *Dama dama*, European moose *Alces alces* and muntjac *Muntiacus reevesi*) the model framework performed well. During the last year, the consortium ENETWILD has collected data on hunting yield (HY) for wild ruminants over Europe at different spatial resolution (e.g., hunting ground, municipalities, NUT3) and therefore next step is incorporating and evaluating the performance of HY based spatial models.

The first goal of this report is modelling HY density (HY) of wild ruminant as response variable. We also update models based on occurrence, evaluate the performance of both approaches, and compare outputs. For comparison between the modelling approaches (HY *vs* occurrence), we focused on two Cervidae species of wide distribution in Europe, namely red deer and roe deer. Models based on HY were statistically downscaled to make predictions to 10x10km squares. We initiate this phase (modelling wild ruminants HY and occurrence for comparison) using these deer species because data availability covers well the continent and lessons learnt on modelling and comparison between models can be applied to further steps, including the remaining wild ungulate species.

Data: We used HY data compiled for the period 2015-2020 from records submitted to ENETWILD Data Model, extracted on 15/12/2021, and occurrences available from the past 10 years (since 2011) through the Global Biodiversity Information Facility (GBIF) extracted on 27/10/2021³ together with latest records from *iMammalia* extracted on 30/11/2021.

Modelling: Like models based on HY, the response variable was the maximum number of wild ruminants (red deer and roe deer) annually hunted in 2015-2020 hunting seasons divided by the area (km²) of the corresponding administrative unit (HY density). We could not allow more flexibility in predictor responses as in previous reports by conducting one model per bioregion (see Glossary) as there are not enough data for modelling in some bioregions. Therefore, we calibrated one model, including eco-geographical variables as predictors. Models based on occurrence used hSDM.ZIB function applying default settings. We considered explanatory variables describing climate, land cover, topography, and human disturbance as in previous ENETWILD reports. Model projections were limited to exclude regions whose environmental conditions were deemed insufficiently represented by the training dataset.

Results and discussion: As for HY models, the general calibration plot for red deer model showed a good predictive performance for the Eastern bioregion, while that of roe deer presented adequate predictive performance at Eastern and Western bioregions. As for the interpretation of spatial patterns of predicted HY, we should be cautious since we could not fit one model per bioregion. The highest density predictions for red deer were reached at the Iberian Peninsula. However, the abundance distribution pattern was widely scattered over all Europe. This is expected for a widely distributed species which shows high ecological plasticity. We found high predicted red deer abundance for the Northern bioregion, where lower abundance and more restricted distribution is expected. Concerning roe deer HY model, the highest abundance is found in Atlantic and Eastern Europe, progressively decreasing towards Northern Mediterranean bioregions. Overall, calibration plot performed better than for red deer, and similarly, low correspondence was found in the Northern region, which may also be due to the low availability of data in this bioregion. Despite overprediction in several calibration plots per species and bioregion, we evidenced linearity in most cases, indicating that the models adequately showed

³ <https://doi.org/10.15468/dl.up2acp>

the spatial pattern. In relation to occurrence data models, overall, model performance using our revised approach was mixed with most species showing similarly moderate predictive accuracy (AUC statistics above or close to 0.7). As for the validation of suitability on HY, red deer showed little variation of observed HY independently of the value of predictive suitability. On the contrary, roe deer showed a good pattern, which means that suitability data are in consonance with observed HY. This may indicate that suitability has potential to be used as a proxy for abundance in roe deer. Although this is a first step and our spatial models will improve in subsequent modelling, overall, our results, except by the Northern bioregion, are consistent with the expected abundance distribution of red deer and roe deer. They are the most common ungulate species in Europe (together with wild boar) and can occupy a diversity of habitats, including deciduous and coniferous continental forests, Mediterranean scrublands, agricultural plains (the latter mainly applies to roe deer), but also high latitudes and altitudes, where harsh winter occur. However, roe deer is less ubiquitous in Mediterranean areas, finding its optimal in Atlantic habitats, and this is well captured by our model.

Conclusions and next steps:

Hunting yield density data

Model projections showed reliable patterns where good quality data was provided, while worst predictions are found in nearer countries/bioregions. This could be due to “over-informing” the model by high amount of data at high resolution in some countries, while there is scarce information and/or poor data quality (spatial resolution) in other countries, which does not have the same weight when modelling. Two approximations to be explored for next models, once more and better data will be available, are: (i) modelling HY per bioregion providing more flexibility to the models, even if data projection is done at lower resolution scales, and (ii), modelling HY by accounting the fact that certain countries provide most data, to avoid that these areas overinforming the model.

Occurrence data model

In isolation, occurrence models produce reasonable predictions with most achieving an AUC of 0.7 or above. However, predictions for the most ubiquitous, easily observable species (i.e., roe deer), were not so accurate. This may be a consequence of the evaluation process itself but may equally be indicative of other issues. Broad target groups currently used to quantify survey effort may need to be refined, particularly for common widespread species. Identification of “true” absence for the purposes of occurrence model testing needs to be reviewed and potentially modified to mimic the process simulated in model fitting. To mitigate difficulties with model evaluation, rather than applying relatively “opaque” principal component analysis (PCA) to arrange and reduce variables, we could instead consider raw ecological variables like HY models; more easily allowing verification against existing knowledge. Next steps for data acquisition and occurrence data modelling are: (i) review target group definitions for each species, (ii) revise definitions of “true” absence for model testing for better parity with fitting, and (iii) either replace principal component analysis with variance inflation factor (VIF) analysis to remove co-correlates and model calibration for variable selection or develop post-model analysis to recover environmental dependencies.

Table of contents

Abstract	1
Summary	2
1. Introduction	6
1.1. Background and Terms of Reference as provided by the requestor	6
1.2. Scope of the report	6
1.3. Environmental variables and other predictors	6
2. Data	8
2.1. Study area	8
2.2. Data collection	9
2.2.1. Red deer and roe deer hunting yield data	9
2.2.2. Wild ruminant occurrence data	11
3. Methodology	12
3.1. Red deer and roe deer: models based on hunting yield data	12
3.2. Models based on occurrence data	15
3.3. Comparison of models based on hunting yield and occurrence data for red deer and roe deer	16
4. Results and discussion	17
4.1. Red deer and roe deer: Model based on hunting yield data	17
4.2. Wild ruminant occurrence data models	21
4.3. Comparison of models based on hunting yield and occurrence data for red deer and roe deer	24
4.4. Conclusions and further steps	25
References	26
Glossary	29

1. Introduction

1.1. Background and Terms of Reference as provided by the requestor

This contract was awarded by EFSA to Universidad de Castilla-La Mancha, contract title: Wildlife: collecting and sharing data on wildlife populations, transmitting animal disease agents, contract number: OC/EFSA/ALPHA/2016/01 – 01.

The terms of reference for the present report (specific contract 8) were to update the occurrence and to parameterize hunting yield-based data models for wild ungulates at European scale. A final report is due by December 2021.

1.2. Scope of the report

The ENETWILD consortium (www.enetwild.com) implemented an EFSA funded project whose main objective has been the collection of information regarding the geographical distribution and abundance of wild boar and other wild ungulates throughout Europe to subsequently create geospatial tools to be used in further risk assessment of diseases, such as African swine fever (ASF) in the case of wild boar. In August 2021 (<https://www.efsa.europa.eu/en/supporting/pub/en-6825>) ENETWILD proposed a generic model framework to predict habitat suitability and likely occurrence for wild ruminant species using opportunistic presence data (occurrence records for wild ungulate species from the Global Biodiversity Information Facility). Across all wild ungulate species (roe deer *Capreolus capreolus*, red deer *Cervus elaphus*, fallow deer *Dama dama*, European moose *Alces alces* and muntjac *Muntiacus reevesi*) the model framework performed well. During the last year, the ENETWILD consortium has collected data on hunting yield (HY) for wild ungulates over Europe at different spatial resolution (e.g., hunting ground, municipalities, NUT3) and therefore the next step is incorporating and evaluating the performance of HY based spatial models.

The goals of this report are modelling the HY densities (HY) of wild ruminants as response variables, to update models based on occurrence, evaluating the performance of both approaches, and comparing outputs. For comparison between modelling approaches (HY vs occurrence), we focused on two Cervidae species of wide distribution in Europe, red deer, and roe deer because data availability covers well the continent and lessons learnt when modelling and comparing HY and occurrence-based models can be applied to other wild ruminant species in further steps.

1.3. Environmental variables and other predictors

According to previous reports (e.g. ENETWILD consortium et al., 2021) we selected environmental variables closely related to wild ungulate distribution describing topography, climate, land cover and human density (Table 1).

Bioclimatic variables and sun radiation were obtained from the Worldclim 2 project database (<https://worldclim.org/version2>). Land use data was downloaded from ESA/CCI-LC project, version v2.0.7 (2015) (<https://www.esa-landcover-cci.org/?q=node/158>). Mean altitude was extracted from the USGS Space Shuttle Radar Topography Mission (SRTM) GL30 (<https://lta.cr.usgs.gov/SRTM1Arc>) and snow cover was obtained from MODIS/Terra Snow Cover project (Monthly L3 Global 0.05Deg CMG, Version 6; <https://nsidc.org/data/MOD10CM>). Human footprint index was provided by The Last of the Wild Project version 2 (<http://sedac.ciesin.columbia.edu/data/collection/wildareas-v2>), while vegetation growing period was obtained from the Agro-Ecological Zones project (FAO; <http://www.appsolutelydigital.com/DataPrimer/part154.html>). The bioclimatic regionalization described in previous reports (ENETWILD consortium et al., 2019a) was maintained for the study

area. According to expert evaluations, in earlier reports some wrong predictions of wild boar abundance were in *Eucalyptus* spp. plantations mainly in West Europe. Those plantations are often considered like forests by telemetry-derived cartographic variables, and suitability indexes calculated for those areas can be misleading. For this reason, in the HY models we considered as predictor the percentage of *Eucalyptus* spp. as dominant species obtained from Brus et al. (2011) (European Forest Institute <https://www.efi.int/knowledge/maps/treespecies>). Raster predictor layers and grid polygons were managed using QGIS 3.4 and *rgdal* R packages (Bivand et al., 2006).

Table 1: Variables used to model (i) the spatial pattern of wild ruminant abundance and (ii) distribution based on hunting yield and occurrence data, respectively.

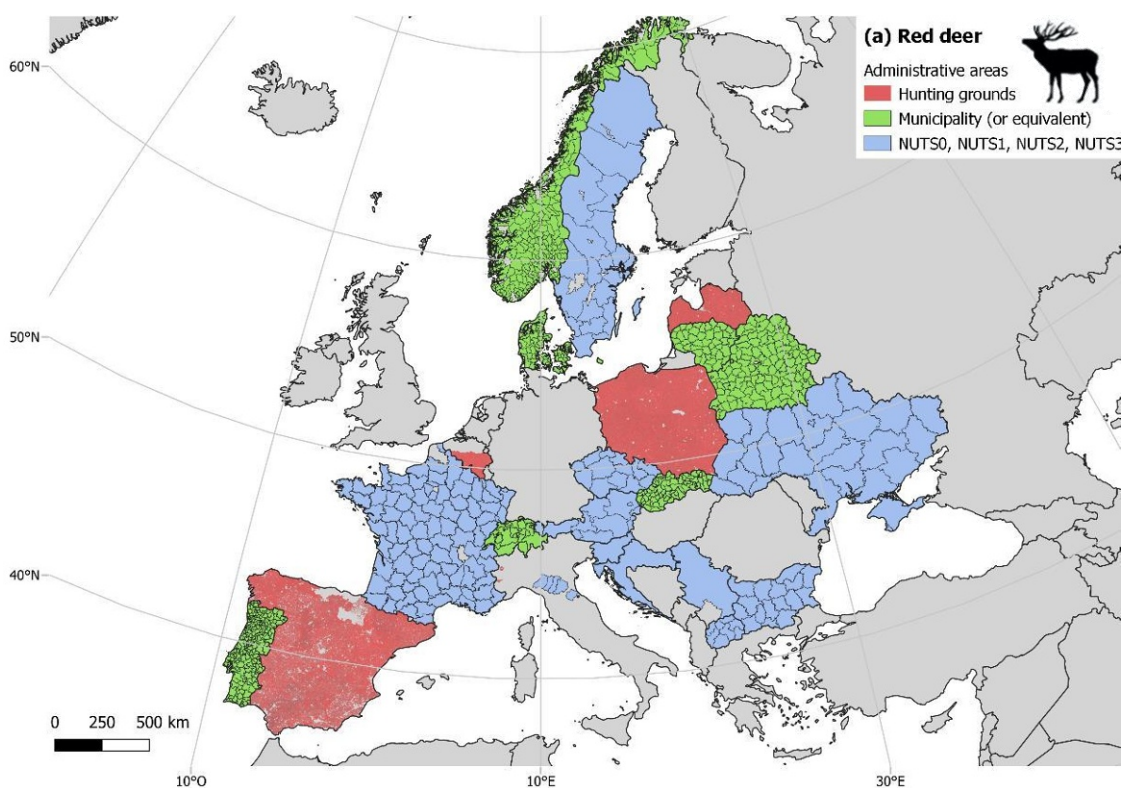
Code	Variable description	Code	Variable description
BIO1	Annual mean temperature	lc_10	Cropland, rainfed
BIO2	Mean diurnal range (mean of monthly (max temp - min temp))	lc_11	Herbaceous cover
BIO3	Isothermality (BIO2/BIO7) (x 100)	lc_12	Tree or shrub cover
BIO4	Temperature seasonality (SD x 100)	lc_20	Cropland, irrigated or post- flooding
BIO5	Max temperature of warmest month	lc_30	Mosaic cropland (>50%) / natural vegetation (tree, shrub, herbaceous cover) (<50%)
BIO6	Min temperature of coldest month	lc_40	Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%) / cropland (<50%)
BIO7	Temperature annual range (BIO5-BIO6)	lc_60	Tree cover, broad-leaved, deciduous, closed to open (>15%)
BIO8	Mean temperature of the Wettest Quarter	lc_61	Tree cover, broad-leaved, deciduous, closed (>40%)
BIO9	Mean temperature of the Driest Quarter	lc_70	Tree cover, needle leaved, evergreen, closed to open (>15%)
BIO10	Mean temperature of warmest quarter	lc_71	Tree cover, needle leaved, evergreen, closed (>40%)
BIO11	Mean temperature of coldest quarter	lc_80	Tree cover, needle leaved, deciduous, closed to open (>15%)
BIO12	Annual precipitation	lc_90	Tree cover, mixed leaf type (broadleaved and needle leaved)
BIO13	Precipitation of wettest month	lc_100	Mosaic tree and shrub (>50%) / herbaceous cover (<50%)
BIO14	Precipitation of driest month	lc_110	Mosaic herbaceous cover (>50%) / tree and shrub (<50%)
BIO15	Precipitation seasonality (coefficient of variation)	lc_120	Shrubland
BIO16	Precipitation of wettest quarter	lc_122	Deciduous shrubland
BIO17	Precipitation of driest quarter	lc_130	Grassland
BIO18	Precipitation of Warmest Quarter	lc_140	Lichens and mosses
BIO19	Precipitation of Coldest Quarter	lc_150	Sparse vegetation (tree, shrub, herbaceous cover) (<15%)
GROW	Length of vegetation growing period	lc_152	Sparse shrub (<15%)
SUNRAD	Sun radiation	lc_153	Sparse herbaceous cover (<15%)

SNOW	Snow cover	lc_160	Tree cover, flooded, fresh or brackish water
HFP	Human Footprint Index	lc_180	Shrub or herbaceous cover, flooded, fresh/saline/brackish water
NUT	Administrative level	lc_190	Urban areas
ALT	Mean altitude	lc_200	Bare areas
AREA	Area of sampling unit	lc_201	Consolidated bare areas
Eu	Percentage of <i>Eucalyptus</i> sp.	lc_202	Unconsolidated bare areas
x_scale	Scaled X coordinate of the centroid of the unit area	lc_210	Water bodies
y_scale	Scaled Y coordinate of the centroid of the unit area	lc_220	Permanent snow and ice

2. Data

2.1. Study area

The study area is the same as the previous report (ENETWILD consortium et al., 2021). It includes all countries in mainland Europe with the Ural Mountains as the eastern limit (Figure 1), spans 11,019,700 km² (110,197 10x10 km and 2,787,877 2x2 km grid cells) and includes Mediterranean islands, the UK, and Ireland.



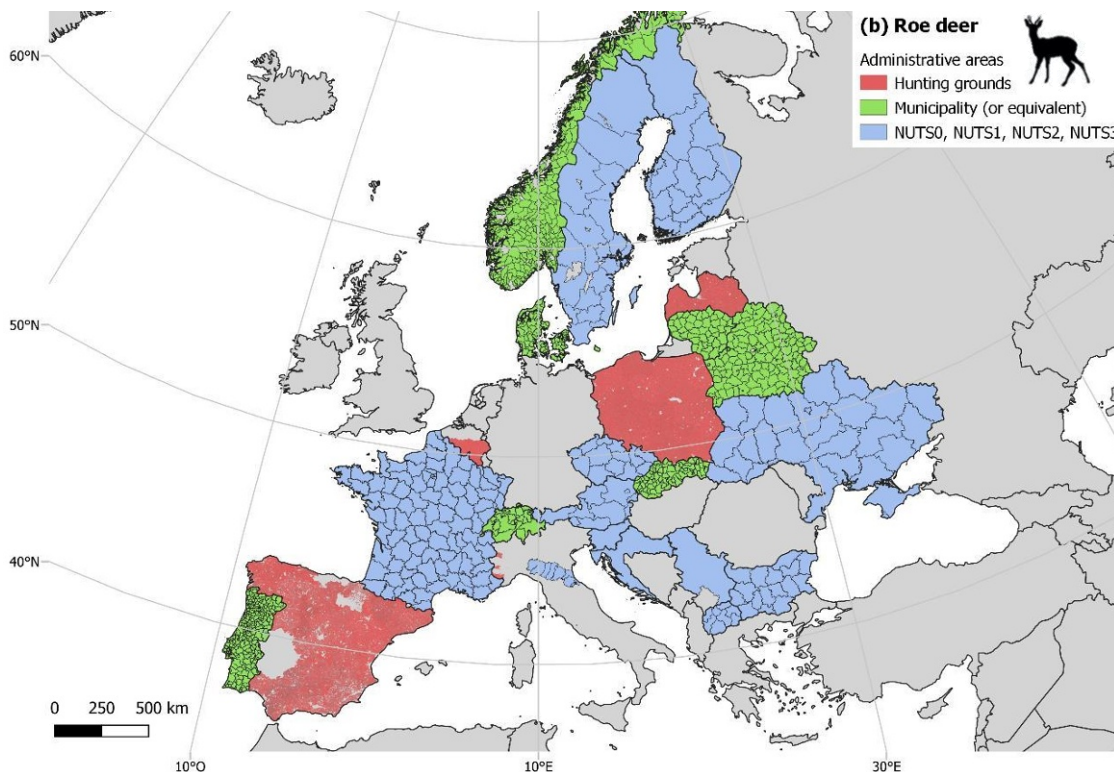


Figure 1: Map showing the extent of the study area. Different colours show the different administrative area levels used (from lowest level 'hunting ground' to highest level 'NUTS 0, NUTS 1, NUTS 2 or NUTS 3') in the models for (a) red deer and (b) roe deer hunting yield.

2.2. Data collection

2.2.1. Red deer and roe deer hunting yield data

For this report hunting yield data were incorporated for modelling from the ENETWILD data collection model downloaded (Figure 1, Table 2). We used 36,247 territorial units for red deer, of which 35,218 correspond to hunting grounds, 839 to municipality level (or equivalent) and 190 to NUT3/NUT2/NUT1/NUT0 (see Figure 1). It must be noticed that Finland did not report data for red deer as it is a non-hunted species. Some fewer territorial units were available for roe deer (33,255), of which 32,041 corresponded to hunting grounds, 1001 to municipality level (or equivalent) and 213 to NUT3/NUT2/NUT1/NUT0 (see Figure 1).

Like previous reports focusing on wild boar, we focused on hunting yield records from 2015 to 2020 hunting seasons. The maximum numbers of red deer and roe deer hunted within the study period per territorial unit were considered for modelling. Data reported by Sweden for 2019 could not have been considered as they correspond to municipality level while previous data of Sweden corresponded to lower spatial resolution (NUTS 3), so the maximum number of hunted deer cannot be assessed at the greater spatial resolution. The response variable was obtained by dividing the maximum number of hunted animals by area (km^2), i.e., we modelled hunting yield density (HY) of red deer and roe deer. We removed "0" records and transformed density data multiplying their values by 100 for modelling purposes (to have integer response variable for the negative binomial models).

Table 2. Number of territorial units of data collection by country.

Country	Territorial units	
	Red deer	Roe deer
Austria	9	9
Belarus	118	118
Belgium	62	62
Bulgaria	16	16
Czech Republic	13	13
Croatia	1	1
Denmark	98	98
Finland	-	15
France	88	93
Italy	23	62
Latvia	449	449
Lithuania	46	46
North Macedonia	8	8
Norway	356	356
Poland	4974	4974
Portugal	278	278
Serbia	1	1
Slovakia	79	79
Slovenia	1	1
Spain	29715	26504
Sweden	21	21
Switzerland	26	26
Ukraine	27	27

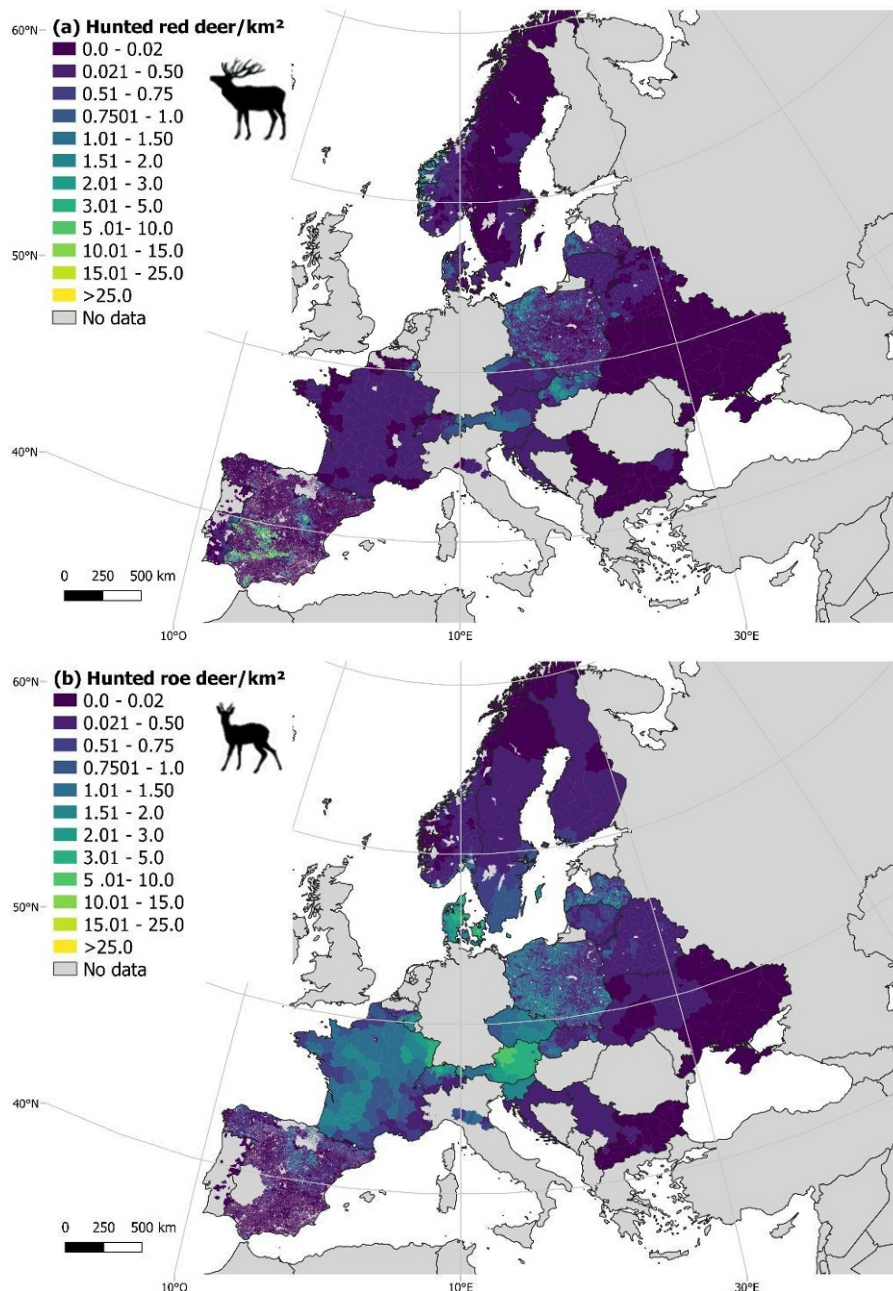


Figure 2: Maximum hunting yield density data (2015-2020 hunting seasons) used for parameterizing the (a) red deer, and (b) roe deer models.

2.2.2. Wild ruminant occurrence data

For modelling species habitat suitability (probability of presence), we obtained occurrence records (presences) from the past 10 years (2011-2020) from datasets hosted on the Global Biodiversity Information Facility (GBIF; extracted on 27/10/2021 <https://doi.org/10.15468/dl.up2acp>) and combined these with records from our own *MammalNet* data collection project (*iMammalia*; ENETWILD consortium et al., 2018b). Any records without an exact sighting date, taxonomic

description to species level and coordinate accuracy equivalent to or better than that required for modelling on a 2x2 km raster grid were excluded.

As in previous reports (ENETWILD consortium et al., 2019b; 2021), we estimated survey effort using a target group approach (Phillips et al., 2009), aggregating sightings across multiple species considered “similar” to the focal species which we wanted to model. In this case we defined target groups based on most common recording method (Ranc et al., 2017): large terrestrial mammals (visual); small terrestrial mammals (trapping); bats (acoustic); and, riverine. To ensure any datasets included in our analysis complied with this idea of multi-species recording we assessed each separately excluding any which contained observations of fewer than 10 different species or 2 taxonomic orders. For each species, we converted the remaining data into a binomial description of occurrence at distinct study sites (2x2 km grid cells) with successes defined as the number of visits (unique dates across all datasets) reporting the focal species and trials (survey effort) as the number of visits reporting any target group species.

3. Methodology

3.1. Red deer and roe deer: models based on hunting yield data

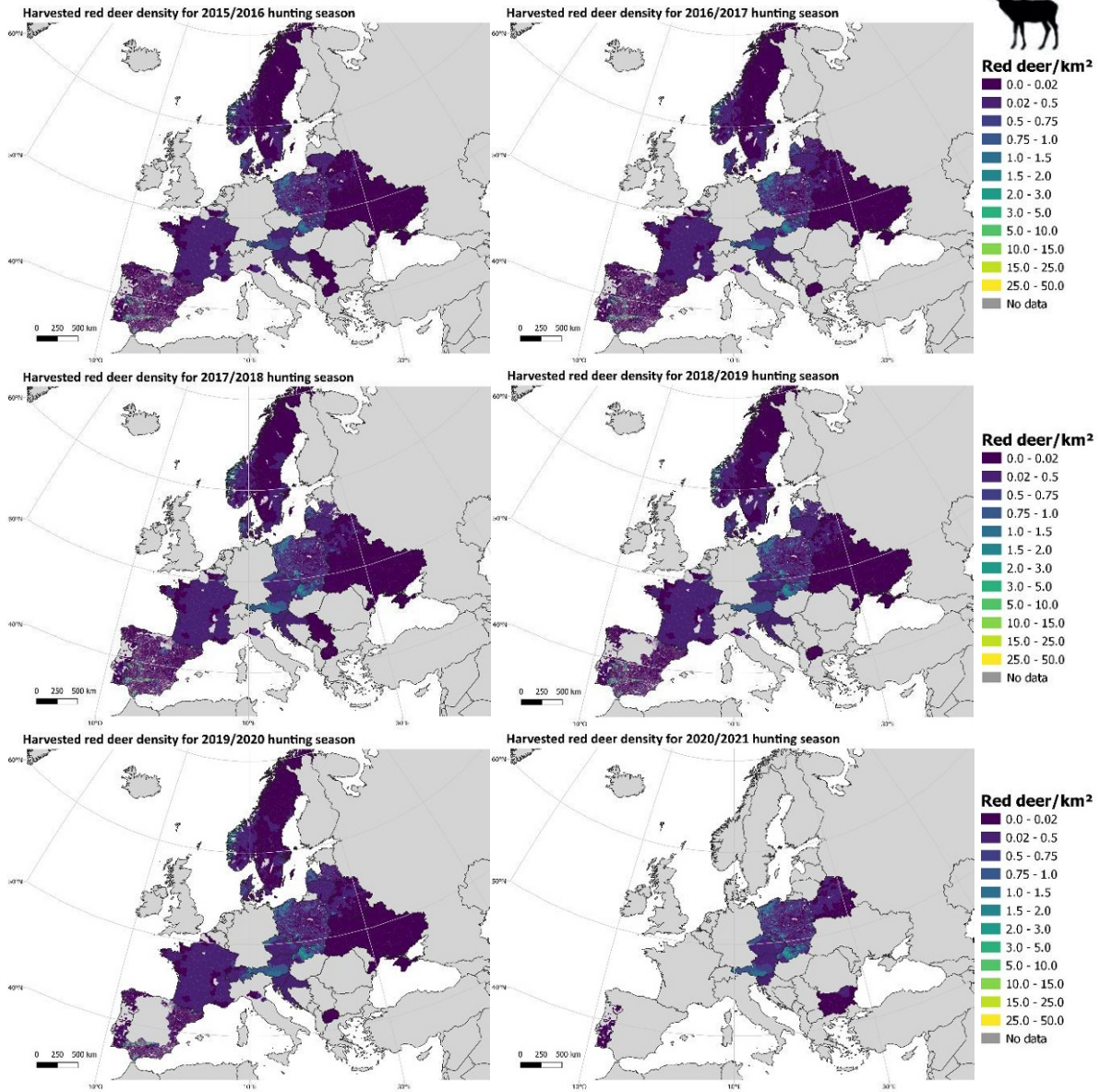
3.1.1. Temporal dataset

It is challenging to correctly manage databases when data are collected at different temporal and spatial resolutions, as shown by the red and roe deer data. The gaps in data availability increased if data collected were separated by years (hunting seasons; Figure 3, Table 3). So that, to reliably analyse temporal data more information is needed.

Table 3. Number of territorial units from 2015/2016 to 2020/2021 hunting seasons.

Hunting season	2015/ 2016	2016/ 2017	2017/ 2018	2018/ 2019	2019/ 2020	2020/ 2021
Red deer	35,699	36,111	36,029	29,690	20,144	5,298
Roe deer	31,815	32,281	32,865	26,495	20,533	5,249

(a)



(b)

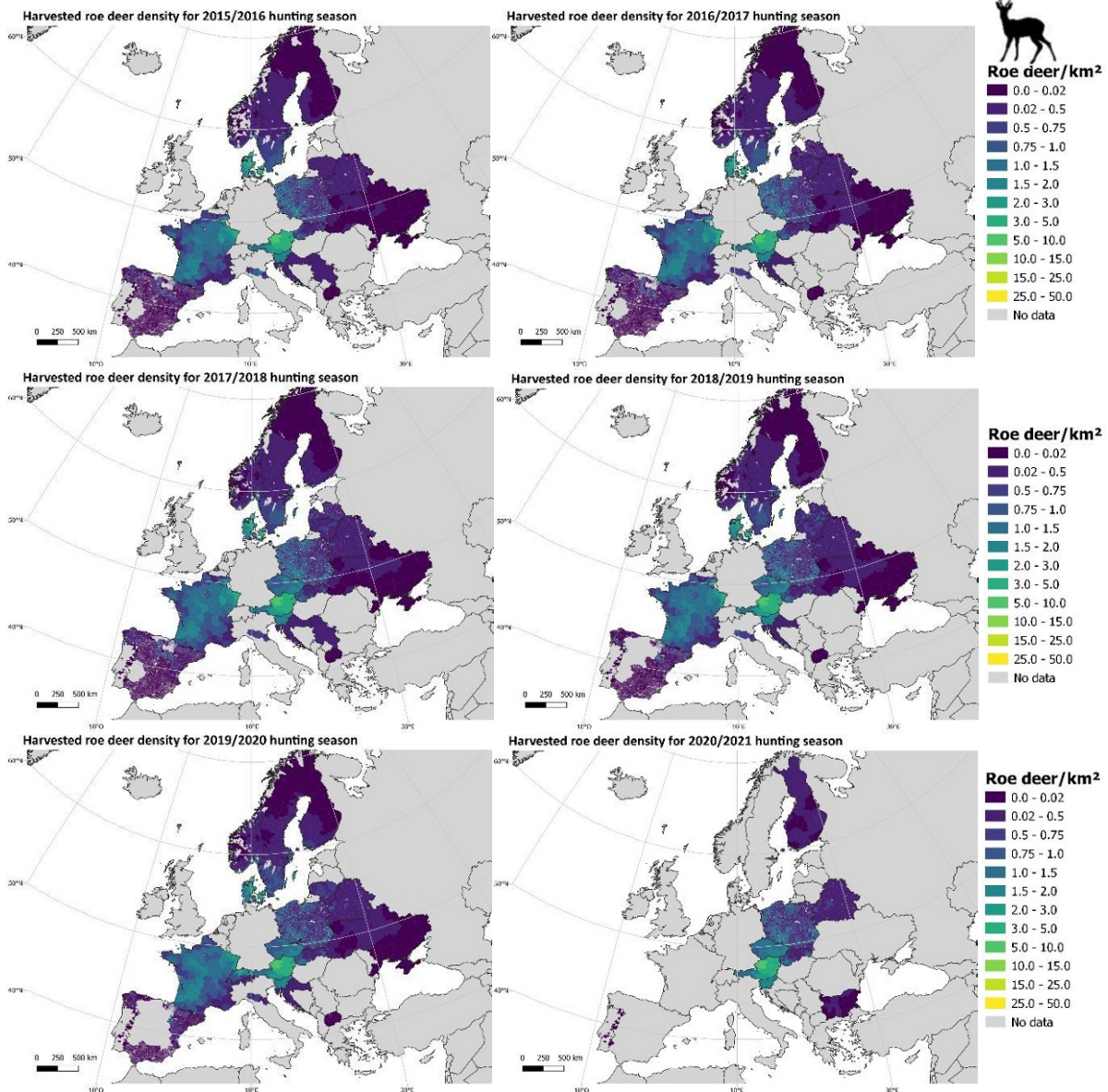


Figure 3. Available data for (a) red deer and (b) roe deer hunting yields (hunted animals per km²) for Europe from 2015/2016 to 2020/2021 hunting seasons.

3.1.2. Modelling procedure for hunting yield data

The response variables for modelling were red deer and roe hunting yield density across Europe, (maximum number of individuals annually hunted within 2015-2020 hunting seasons, divided by unit area in km²; hereafter HY).

We could not allow more flexibility in predictor responses as in previous reports by conducting one model per bioregion as there were not enough data for modelling in some bioregions (Table 4; see also Acevedo et al., 2014; Pittiglio et al., 2018; ENETWILD consortium et al., 2020a). Therefore, we calibrated one model per species, including eco-geographical variables as

predictors. The following steps were the same as in previous reports. Eco-geographical predictors more relevant in explaining HY were determined using generalized linear models (negative binomial distribution and logarithmic link function; Cameron and Trivedi 2013). Multicollinearity among predictors was assessed using Variance Inflation Factor (VIF); predictors with VIF values above 2 were removed (Zuur et al., 2010).

All models were trained using an 80% random sample of the data (training dataset) and model predictions were validated against the remaining 20% of the data (validation dataset). The final models were obtained using forwards-backwards stepwise procedure based on Akaike Information Criteria (AIC; Akaike 1974).

After modelling, calibration plots were developed to assess the predictive performance of the model. This was carried out by plotting the mean observed HY in each interval (defined from percentiles) of the predicted HY on the validation dataset, and thus perfect predictions should lie along the identity line (Pearce and Ferrier, 2001), where linearity of the relationship informs about the reliability of the predicted pattern. Moreover, we divided the validation data into the four bioregions to assess if the model fit differed among bioregions and used the calibration dataset with and without 0 density values.

Model output was statistically downscaled to make predictions at 10x10km using EAA grid (<https://www.eea.europa.eu/data-and-maps/data/eea-reference-grids-2>).

Table 4. Number of territorial units per bioregion and administrative area. Data are shown for both red and roe deer.

Bioregion	Administrative area	Red deer	Roe deer
Northern	Hunting ground	0	0
	Municipality (or equivalent)	107	152
	NUTS0, NUTS1, NUTS2, NUTS3	7	21
Southern	Hunting ground	27,875	24,119
	Municipality (or equivalent)	232	232
	NUTS0, NUTS1, NUTS2, NUTS3	6	6
Eastern	Hunting ground	5423	5423
	Municipality (or equivalent)	295	295
	NUTS0, NUTS1, NUTS2, NUTS3	76	77
Western	Hunting ground	1920	2499
	Municipality (or equivalent)	205	322
	NUTS0, NUTS1, NUTS2, NUTS3	101	109

3.2. Models based on occurrence data

To estimate distributions for each species of interest we applied a refined version of the model framework outlined in our previous report (ENETWILD consortium et al., 2021). To account for various biases and ensure that model assumptions were observed, we performed additional filtering of survey sites. To ensure stationarity (distributions not limited by dispersal), we excluded

sites beyond the estimated “stable” species range based on a combination of expert-derived maps (Burgin et al., 2020; MDD, 2020; IUCN 2021; Wilson et al., 2021) and maps derived from occurrence data outlining regions with consistent sightings (Maes et al., 2015). To ensure sampling independence we reduced spatial clustering by thinning sites. For the latter, we derived an optimal site spacing to ensure both environmental (Di Cola et al., 2017) and ecological (Aiello-Lammens et al., 2015) independence (see Table 7 for details of species-specific spacing), which we used to produce stochastic sub-samples of sites (10 per model run).

For each thinned sample, we modelled the probability of occurrence across sites using the “hSDM.ZIB” function of the “hSDM” package (Vieilledent et al., 2014) in R statistical software (R Development Core Team 2018) applying default settings. This function used our binomial dataset within a hierarchical Bayesian framework integrating two processes: (i) an ecological process, represented by a Bernoulli distribution, describing species presence or absence due to environmental suitability; (ii) an observation process, represented by a binomial distribution, which considers the fact that detection of the species is imperfect (i.e., the probability of detecting the species is less than 1) (MacKenzie et al., 2002; Latimer et al., 2006).

To model the ecological process, we considered explanatory variables describing climate, land cover, topography, and human disturbance as in previous ENETWILD reports (Table 1). Prior to model fitting we used a principal component analysis, rearranging this set of variables to minimise co-correlation, and applied a scaled average-eigenvalue test (Kaiser 1960) to remove any redundancy (ENETWILD consortium et al., 2019b). For the observation process we only considered a constant to reflect that the number of trials was derived from other presence records whose detectability within any given cell was likely to be similarly affected by the environmental condition, therefore detectability was represented in relative terms compared to that of other species (Croft et al., 2019). For all model parameters (coefficients) in both ecological and observation processes we used default uninformative Normal priors with a mean of zero and large variance of $1e+06$ providing a relatively flat distribution. Based on a MESS analysis (Elith et al., 2010), model projections were limited to exclude regions whose environmental conditions were deemed insufficiently represented by the training dataset so as not to produce unreliable prediction. Model projections across all repetitions were combined to produce a single mean output.

We assessed model performance using 4-fold cross validation, systematically reserving a proportion of data (25%) for testing and applying the model fitting process described above to the remainder to predict suitability and observability. For each fold, we used predicted observability to identify “true” absences within the testing data. Together with the known presences we further sub-sampled the testing data to account for spatial sorting biases (Hijmans, 2012) which are known to inflate evaluation metrics (Neftalí and Barbosa, 2021). Using this dataset, we computed several common metrics for predictive accuracy (see Glossary for further details): AUC (area under curve statistic, calibrated against a null model; Hijmans, 2012); TPR (True positive rate - Sensitivity); TNR (True negative rate - Specificity); and TSS (True skill statistic). Finally, we produced mean statistics across all folds.

3.3. Comparison of models based on hunting yield and occurrence data for red deer and roe deer

To validate suitability on HY, calibration plots were developed to assess the relationship between predictive model suitability with observed HY of red deer and roe deer. We calculated the mean suitability for each territorial unit (hunting grounds, municipalities, or NUTS) and defined 9 intervals (defined from percentiles) from it. Then it was calculated the observed HY, and their relation was plotted to assess the relationship pattern for each species, respectively.

4. Results and discussion

4.1. Red deer and roe deer: Model based on hunting yield data

After VIF analyses, we considered for modelling the variables listed in Table 5 and after a stepwise procedure the final models for red deer and roe deer were obtained (see Table 6).

Table 5: Variables selection after VIF analyses for modelling red deer and roe deer abundance (Codes as in Table 1).

	Environmental variables selected for models
Red deer	x_scale+y_scale+bio_4+bio_8+bio_15+alt+snow+Eu+ lc10+lc30+lc40+lc60+lc70+lc90+lc100+lc120+lc130+NUT
Roe deer	x_scale+y_scale+bio_2+bio_4+bio_8+bio_15+alt+snow+Eu+lc10+lc30+lc40+lc60+lc70+lc100+lc120+lc130+NUT

Table 6: Statistical parameters for the final HY models obtained to explain variation in red deer and roe deer in Europe. Codes as in Table 1.

Model	Formula
Red deer	4.848e-01 -1.14e*x_scale -4.559e-03*lc10 +2.262e-02*bio_15 +3.933e-02*lc90 +2.299e-03*bio_4 +1.010e-01*bio_8 +1.408e-02*lc70 +1.876e-02*lc60 +1.976e-02*lc120 +1.626e-02*lc100 +1.691e-02*lc40 +1.474e-02*lc30 -6.027e-02*Eu +2.524e-02*y_scale +7.483e-02*lc130 +7.483e-03*lc130 +4.822e-03*NUT_Municipality -6.529e-01*NUT_0123 +1.054e-04*alt
Roe deer	5.924 -5.0148e-02*bio_2 -4.137e-02snow -9.951e-03*bio_15 +2.683e-01*x_scale +7.079e-03*lc60 + 5.848e03*lc100 -1.435e-01*Eu -1.759e-03*bio_4 +2.108e-01*y_scale +2.015e-02*bio_8 +1.146e-04*alt +2.582e-03*lc130 +1.079e-03*lc10

The two models were then projected at 10x10km grid for the entire study region. HY models' predictions suggested different patterns depending on the study species (Figure 4).

The general calibration plot for red deer HY density model (Figure 5a) showed a good predictive performance for the first bins, although it showed an overfitted pattern for the remaining. The general pattern is very similar to the Southern bioregion specific pattern, probably due to the high amount of data in this region. The Eastern bioregion showed a good predictive performance, however Northern and Western bioregions showed a bad predictive performance.

The general calibration plot for roe deer HY model (Figure 5b) showed an overpredicted pattern for all categories, and it also had a pattern like that of the Southern bioregion, as happened with red deer. An indetermined pattern is found in Northern bioregion, showing under or overprediction depending on the bin. Better predictive performance is found at Eastern and Western bioregions, showing both a good predictive performance.

Overall, despite overprediction in several calibration plots, we demonstrated linearity in most cases, indicating that the models adequately represented the spatial pattern. Moreover, the removal of zeros on the validation dataset had a small effect on the calibration plots of all regions, except the Southern region, which showed a good predictive performance. The two patterns found (all data vs removing 0) could point out areas of potential expansion of the species according to environmental characteristics.

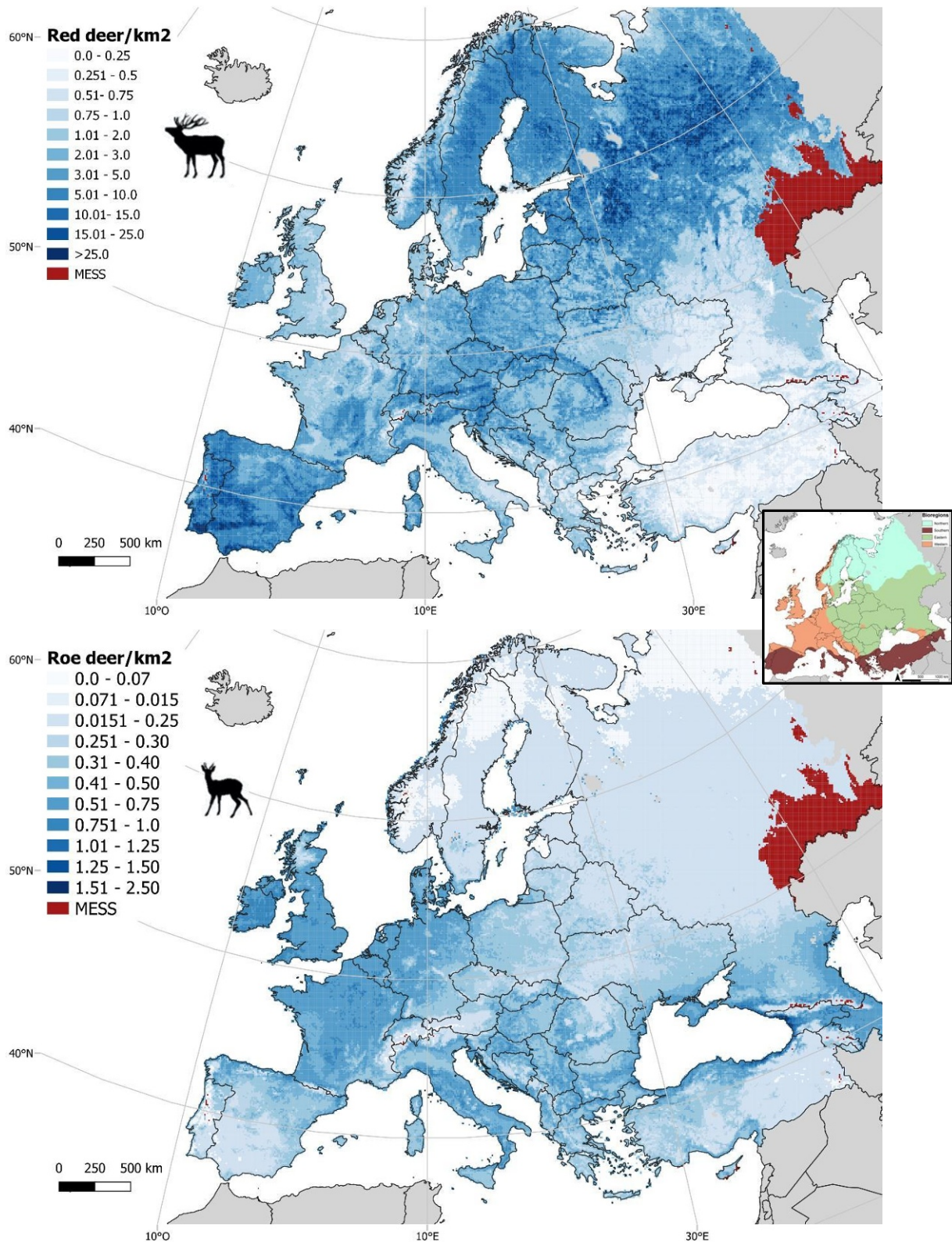
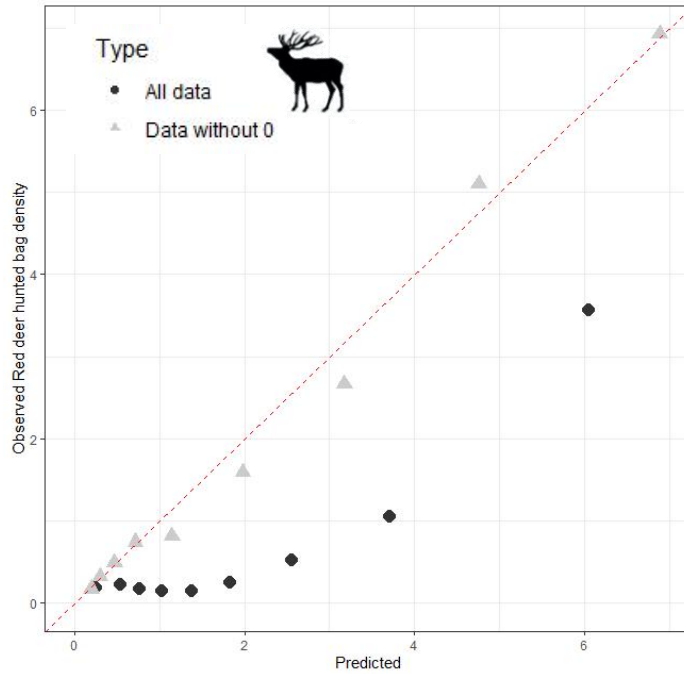
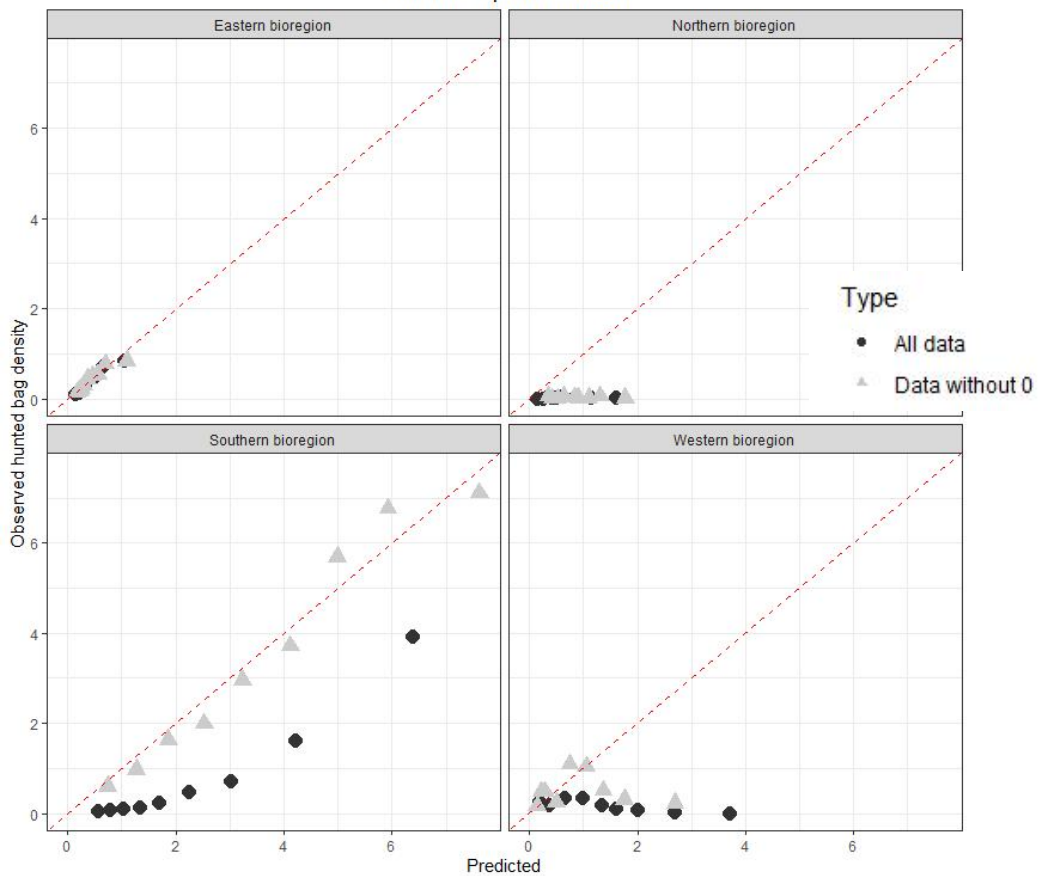


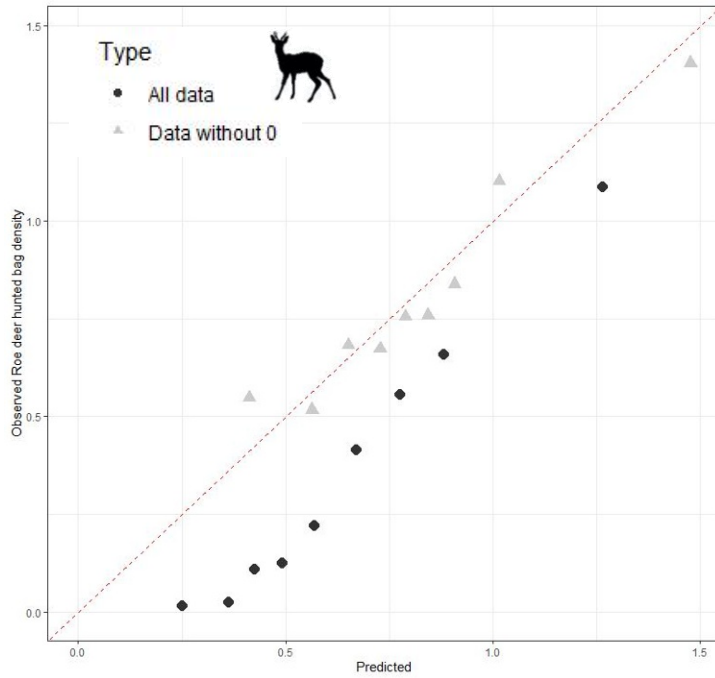
Figure 4: Hunting yield (HY) density (individual hunted per Km²) of red deer (a) and roe deer (b) at 10x10km. Red areas are beyond the environmental domain according to MESS analyses. The window on the right shows the bioregion classification for further discussion purposes.



(a)

Calibration plot Red deer





(b)

Calibration plot Roe deer

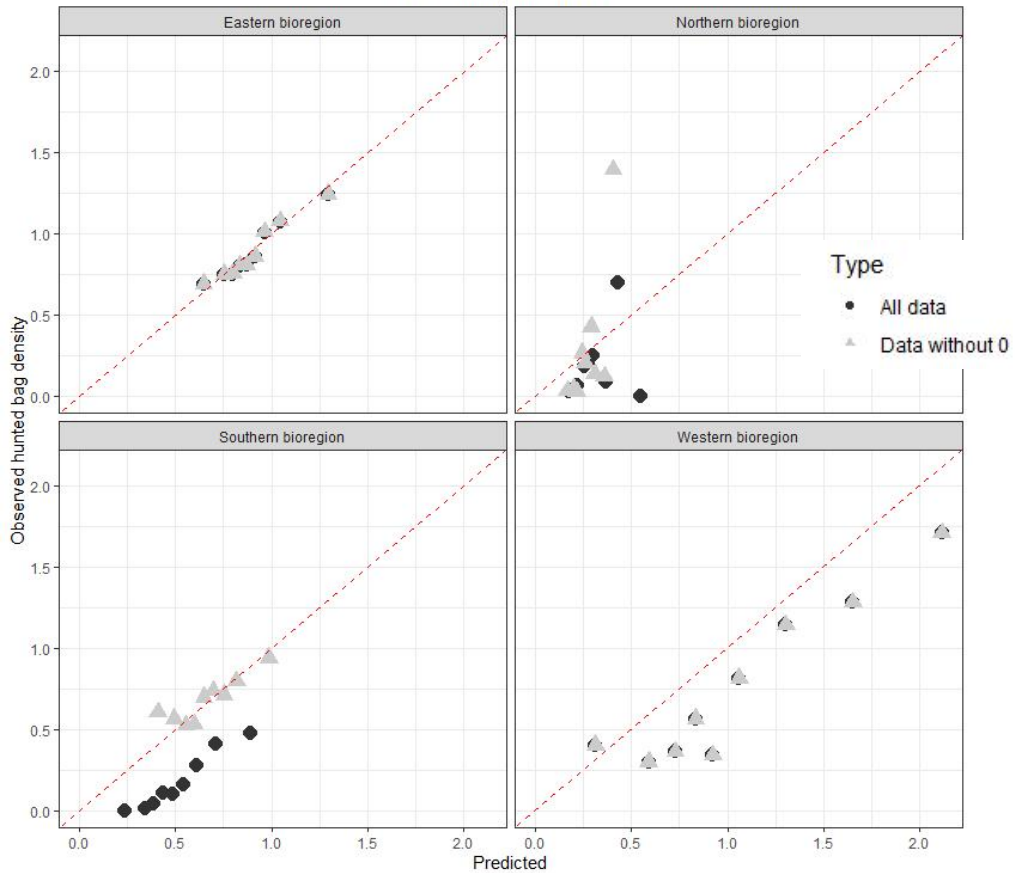


Figure 5: Calibration plot for assessing the predictive performance of (a) red deer HY model, for all Europe and by bioregions and (b) roe deer HY model, again for all Europe and by bioregions (Table 5). Plots show the relationship between the predicted hunting yield densities (HY) and the observed ones on the validation datasets.

As for the interpretation of spatial patterns of abundance (predicted HY) for red deer and roe deer, we exercised caution in this report since we could not allow more flexibility in predictor responses as in previous reports by conducting one model per bioregion as there are not enough data for modelling in some bioregions (Table 4; ENETWILD consortium et al., 2020a). However, this first attempt to model wild ruminant abundance in Europe based on HY is of great value and opens new work lines once more and better distributed data (balanced among bioregions) are available.

The highest HY predictions for red deer were reached at the Iberian Peninsula. However, the abundance distribution pattern was widely scattered over all Europe. This is expected for a widely distributed species which shows high ecological plasticity. The high abundance predicted for the Northern bioregion was especially striking (see Figure 5a), where lower abundance and a more restricted distribution is expected, which may be due to the low availability of data in this bioregion.

Concerning roe deer, the highest abundance is found in temperate Atlantic (including British Isles) and in Eastern Europe, progressively decreasing towards North (Northern region) and South (Mediterranean region), respectively. Overall, the calibration plot performed better than for red deer, and similarly, low correspondence was found in the Northern region (Figure 5), which may also be due to the low data availability in this bioregion.

Although this is a first step and our spatial models will improve in subsequent reports (see next steps below), overall, our results, except by the Northern bioregion, are consistent with the expected abundance distribution of red deer and European roe deer. They are the most common ungulate species in Europe (together with wild boar) and can occupy a diversity of habitat, including deciduous and coniferous continental forests, Mediterranean scrublands, agricultural plains (the latter mainly applies to roe deer), but also high latitudes and altitudes, where harsh winter occur. However, roe deer is less ubiquitous in Mediterranean areas, finding its optimal in Atlantic habitats, and this is well captured by our model.

4.2. Wild ruminant occurrence data models

Mapped predictions from the model showing updated habitat suitability and corresponding occurrence for a selection of the most common wild ruminant species are provided in Figure 6. Evaluation statistics assessing model performance are shown in Table 7. Overall, model performance using our revised approach was mixed with most species showing similarly predictive accuracy with AUC statistics above or close to 0.7. Predictions where AUC remained well below 0.7, notably roe deer, showed lower accuracy in predicting absences which may be linked to high observability (which translates into a lower threshold to declare true absence).

Table 7. Summary of derived model parameters (minimum spacing in km used to thin datasets), distribution statistics (estimated “stable” range size, occupancy within range and potential occupancy across the model extent derived from thresholded model outputs, Figure 6, occurrence, including areas where prediction may be uncertain, in km²), predicted observability (probability) and mean evaluation metrics computed across repetitions for each species (AUC: Area Under the Curve; THD: Optimal suitability threshold defining binary presence/absence which maximises TSS (Liu et al., 2013); SE: Sensitivity; SP: Specificity; TSS: True Skill Statistic).

Species	Sp	Rng	Occ.	Pot.	Obs.	AUC	THD	SE	SP	TSS
<i>A. alces</i>	14	5,793,212	3,221,188	3,401,488	0.21	0.74	0.57	0.72	0.66	0.38
<i>C. capreolus</i>	34	7,168,932	4,876,860	6,415,440	0.37	0.52	0.42	0.88	0.21	0.09
<i>C. elaphus</i>	34	5,664,584	2,400,156	5,376,836	0.28	0.72	0.18	0.66	0.66	0.32
<i>D. dama</i>	34	2,927,220	1,500,640	6,406,480	0.16	0.70	0.32	0.73	0.55	0.28
<i>M. reevesi</i>	10	98,004	67,436	8,153,796	0.10	0.65	0.55	0.85	0.40	0.26

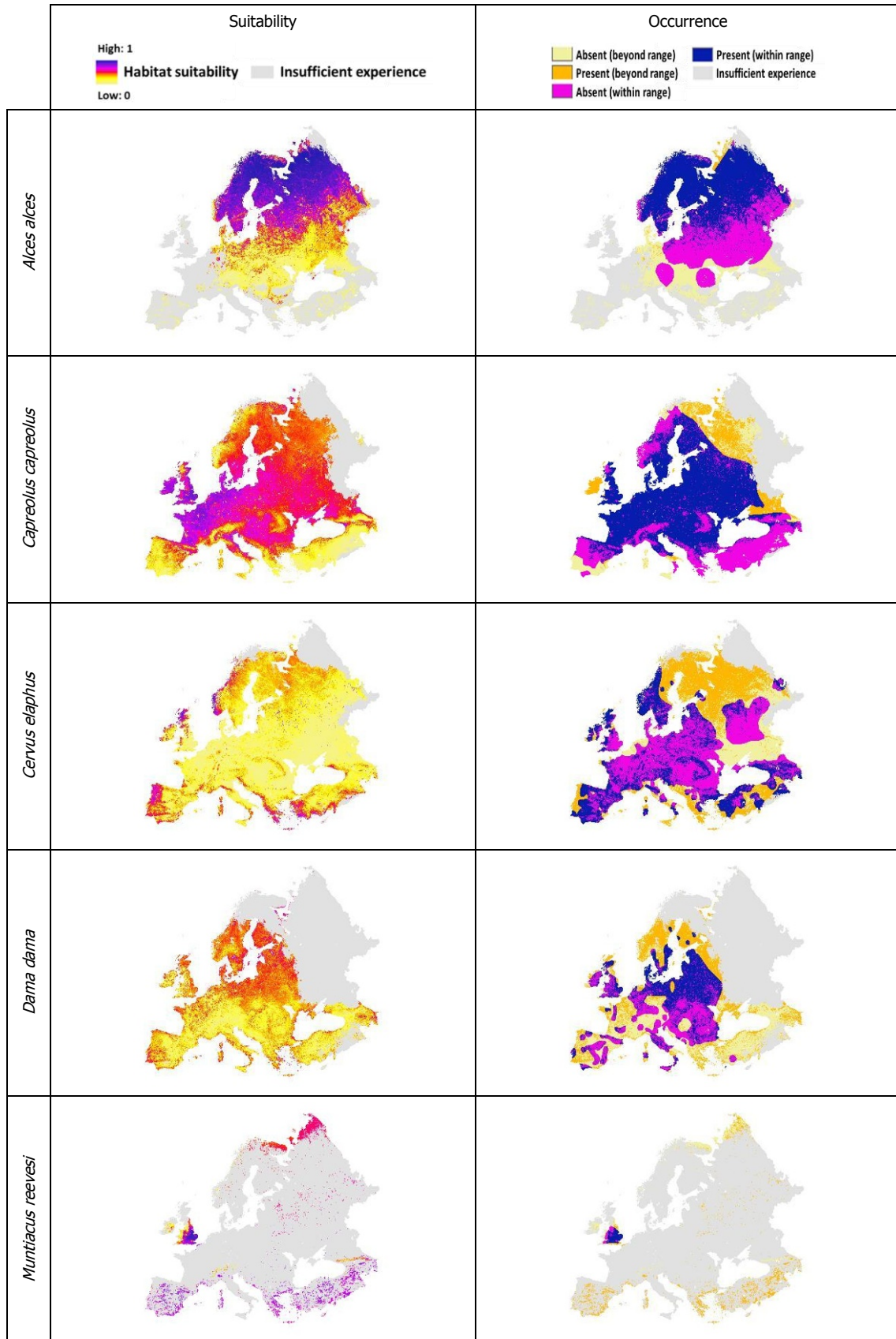


Figure 6: Predictions of wild ruminant occurrence data models. Mapped output showing projected habitat suitability (mean across repetitions ignoring individual exclusions where model transferability may be limited) and threshold occurrence (based on mean habitat suitability and estimated threshold using cross-validation; mess regions based on complete dataset where inference may be unreliable) for each species. The legend “insufficient experience” refers to the lack of information regarding survey effort. In most cases, e.g., *Alces alces* in southern Europe, absence of data is evidence of true absence of a species.

4.3. Comparison of models based on hunting yield and occurrence data for red deer and roe deer

The calibration plot for red deer shows a linear flat pattern (Figure 7), showing little variation of observed HY independently of the suitability value. On the contrary, roe deer showed a good pattern, especially at first suitability intervals, which means that spatial pattern for localities with low suitability (under 0.3) match the pattern displayed from observed HY, while for the last intervals, the slope of the relationship changed.

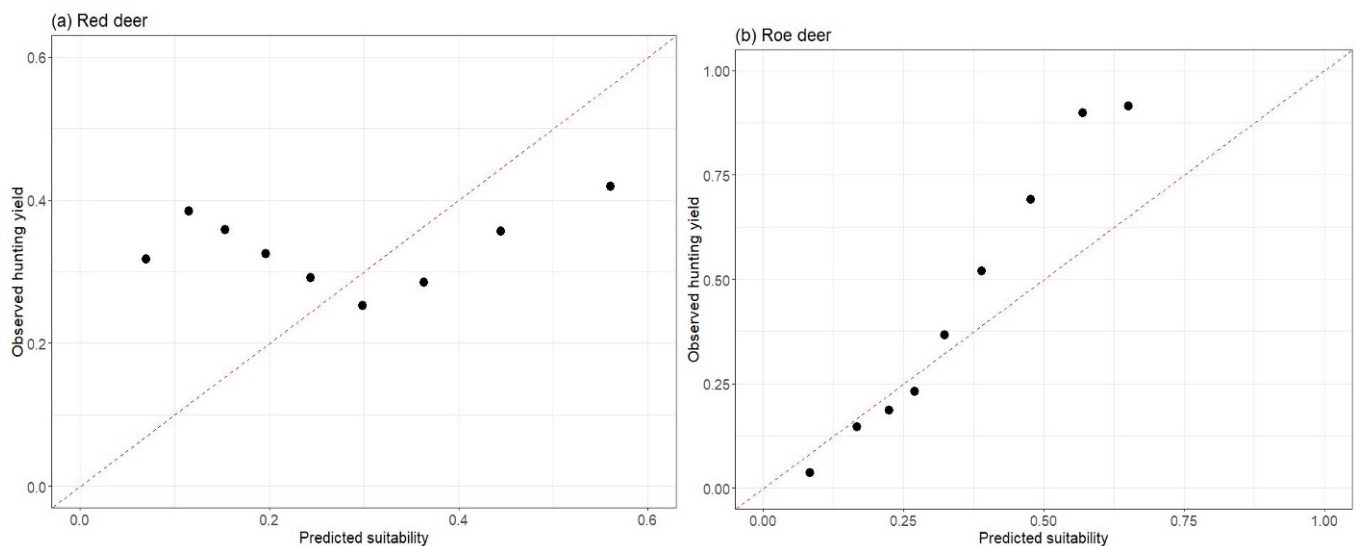


Figure 7: Calibration plots of (a) red deer and (b) roe deer showing the relationship between predicted suitability and observed hunting yields values.

4.4. Conclusions and further steps

Hunting yield density data model

- Model projections showed good patterns where good quality data was provided, while worst predictions are found in neighbouring countries/bioregions where poor quality data were available. This could be due to “over-informing” the model by high amount of data at high resolution in some countries, while there is scarce information and/or data quality (spatial resolution) in other countries, which does not have the same weight when modelling.
- Two approximations to be explored for next models, once more and better data will be available, are:
 - Modelling HY per bioregion providing more flexibility to the models, even if data projection is done at lower resolution scales (if data at higher resolution are still missing or are scarce for some bioregions).
 - Modelling HY by accounting the fact that certain countries provide most data by including random effect factors, to avoid that these areas overinform the model.
- As it has been seen that suitability has a good relationship with roe deer HY models, it could be explored to incorporate it into the HY models to see if it increases the explained variance of the model.

Occurrence data model

- In isolation, occurrence models produce reasonable predictions with most achieving an AUC of 0.7 or above. However, predictions for the most ubiquitous, easily observable species (i.e., roe deer), are not so accurate. This may be a consequence of the evaluation process itself but may equally be indicative of other issues.
- Broad target groups currently used to quantify survey effort may need to be refined, particularly for common widespread species in line with observations made by Ranc et al., (2017).
- Identification of “true” absence for the purposes of occurrence model testing needs to be reviewed and potentially modified to mimic the process simulated in model fitting.
- To mitigate difficulties with model evaluation, rather than applying relatively “opaque” principal component analysis (PCA) to arrange and reduce variables, we could instead consider raw ecological variables like HY models; more easily allowing verification against existing knowledge (ENETWILD consortium et al., 2018a).
- Next steps for data acquisition and modelling are:
 - Review target group definitions for each species (e.g., Croft and Smith, 2019).
 - Revise definitions of “true” absence for model testing for better parity with fitting.
 - Replace principal component analysis with VIF analysis to remove co-correlates and model calibration for variable selection.

Validation of suitability on HY

The comparison of models based on HY and occurrence data indicated that suitability predicted values are in consonance with hunting yields values for roe deer. This may indicate that suitability has potential to be used as a proxy for abundance in this species, but not in red deer. To elucidate the causes of this contrasting result, we must evaluate in the future, once new data are incorporated, how the patterns obtained for the different deer species are affected by specific

predictors. Also, there is need to evaluate the effect of environmental predictors (which explained well the pattern for roe deer) compared to management or anthropic factors (which, we speculate, may be more relevant for red deer). As HY model seemed to behave differently depending on the bioregion, it could be interesting to assess whether there are bioregions where the fit is better.

References

- Acevedo P, Quirós-Fernández F, Casal C, Vicente J, 2014. Spatial distribution of wild boar population abundance: Basic information for spatial epidemiology and wildlife management. *Ecological Indicators* 36: 594-600.
- Aiello-Lammens ME, Boria RA, Radosavljevic A, Vilela B, Anderson RP, 2015. spThin: An R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*, 38: 541-545.
- Akaike H, 1974. A new look at the statistical model identification. *IEEE transactions on automatic control* 19: 716-723.
- Alexander NS, Massei G, Wint W, 2016. The European Distribution of *Sus Scrofa*. Model Outputs from the Project Described within the Poster - Where are All the Boars? An Attempt to Gain a Continental Perspective. *Open Health Data* 4: e1.
- Bivand R, 2006. Implementing spatial data analysis software tools in R. *Geographical Analysis*, 38: 23-40.
- Brus DJ, Hengeveld GM, Walvoort DJJ, Goedhart, PW, Heidema AH, Nabuurs GJ, Gunia K, 2012. Statistical mapping of tree species over Europe. *European Journal of Forest Research*, 131: 145-157.
- Burgin CJ, Wilson DE, Mittermeier RA, Rylands AB, Lacher TE, Sechrest W (Ed.), 2020. Illustrated Checklist of the Mammals of the World. Lynx Edicions. Map of Life. 2021. Mammal range maps digitized from the Illustrated Checklist of the Mammals of the World (Burgin et al., 2020) [Data set]. Map of Life.
- Cameron AC, Trivedi PK, 2013. Regression analysis of count data. Volume 53. Cambridge university press.
- Croft S, Ward AI, Aegerter JN, Smith GC, 2019. Modelling current and potential distributions of mammal species using presence-only data: A case study on British deer. *Ecology and Evolution*, 9: 8724-8735.
- Croft S, Smith GC, 2019. Structuring the unstructured: estimating species-specific absence from multi-species presence data to inform pseudo-absence selection in species distribution models. *bioRxiv*, 656629.
- Di Cola V, Broennimann O, Petitpierre B, Breiner FT, D'Amen M, Randin C, Engler R, Pottier J, Pio D, Dubuis Pellissier L, Mateo RG, Hordijk W, Salamin N, Guisan A, 2017. ecospat: An R package to support spatial analyses and modelling of species niches and distributions. *Ecography*, 40: 774-787.
- Elith J, M. Kearney M, Phillips S, 2010. The art of modelling range-shifting species. *Methods in Ecology and Evolution*, 1:330-342.
- ENETWILD-consortium, Croft S, Smith G, Acevedo P, Vicente J, 2018a. Wild boar in focus: Review of existing models on spatial distribution and density of wild boar and proposal for next steps. *EFSA Supporting Publications* 15:1490E.
- ENETWILD-consortium, Vicente J, Plhal R, Blanco-Aguiar JA, Sange M, Podgórski T, Petrovic K, Scandura M, Cohen Nabeiro A, Body G, Keuling O, Apollonio M, Ferroglio E, Zanet S, Brivio F, Smith GC, Croft S, Acevedo P, Soriguer R, 2018b. Analysis of hunting statistics collection frameworks for wild boar across Europe and proposals for improving the harmonisation of data collection. *EFSA supporting publication*, 15(12), EN-1523. 33 pp.
- ENETWILD-consortium, Acevedo P, Croft S, Smith GC, Blanco-Aguiar JA, Fernandez-Lopez J, Scandura M, Apollonio M, Ferroglio E, Keuling O, Sange M, Zanet S, Brivio F, Podgorski T, Petrovic K, Soriguer R, Vicente J, 2019a. ENETWILD modelling of wild boar distribution

- and abundance: update of occurrence and hunting data-based models. EFSA Supporting Publications, 1: 1674E.
- ENETWILD-consortium, Croft S, Smith G, Acevedo P, Vicente J, 2019b. Wild boar in focus: initial model outputs of wild boar distribution based on occurrence data and identification of priority areas for data collection. EFSA Supporting Publications 16: 1533E.
- ENETWILD-consortium, Acevedo P, Croft S, Smith G, Vicente J, 2019c. ENETWILD modelling of wild boar distribution and abundance: initial model output based on hunting data and update of occurrence-based models. EFSA supporting publication 2019:EN-1629.
- ENETWILD-consortium, Acevedo P, Croft S, Smith GC, Blanco-Aguilar JA, Fernandez-Lopez J, Scandura M, Apollonio M, Ferroglio E, Keuling O, Sange M, Zanet S, Brivio F, Podgorski T, Petrovic K, Soriguer R, Vicente J, 2020a. Update of occurrence and hunting yield-based data models for wild boar at European scale: new approach to handle the bioregion effect. EFSA Supporting Publications, 17(5): 1871E.
- ENETWILD-consortium, Fernandez-Lopez J, Apollonio M, Blanco-Aguilar JA, Brivio F, Croft S, Fanelli A, Fernández-Arias A, Ferroglio E, Keuling O, Levanič T, Plis K, Podgorski T, Pokorny B, Scandura M, Smith GC, Soriguer R, Vicente J, Zanet S, Acevedo P, 2020b. Improving models of wild boar hunting yield distribution: new insights for predictions at fine spatial resolution. EFSA Supporting Publications, 17(12): 1980E.
- ENETWILD-consortium, Acevedo P, Croft S, Smith GC, Blanco-Aguilar JA, Fernandez-Lopez J, Scandura M, Apollonio M, Ferroglio E, Keuling O, Sange M, Zanet S, Brivio F, Podgorski T, Petrovic K, Soriguer R, Vicente J, 2020c. Validation and inference of high-resolution information (downscaling) of ENETWILD abundance model for wild boar. EFSA Supporting Publications, 17(1), 1787E.
- ENETWILD-consortium, Illanas S, Croft S, Smith G C, Fernández-López J, Vicente J, Blanco-Aguilar J A, Pascual-Rico R, Scandura M, Apollonio M, Ferroglio E, Keuling O, Zanet S, Brivio F, Podgorski T, Plis K, Soriguer RC, Acevedo P, 2021. Update of hunting yield-based data models for wild boar and first models based on occurrence for wild ruminants at European scale. EFSA Supporting Publication 2021:EN-6825.
- Hijmans RJ, 2012. Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology*, 93: 679-688.
- IUCN, 2021. The IUCN Red List of Threatened Species. International Union for Conservation of Nature. Accessed October 2021.
- Kaiser HF, 1960. The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20: 141-151.
- Jiménez-Valverde A, Lobo JM, 2007. Threshold criteria for conversion of probability of species presence to either-or presence-absence. *Acta Oecologica*, 31: 361-369.
- Jolliffe IT, 1972. Discarding variables in a principal component analysis. I: Artificial data. *Applied statistics* 160-173.
- Latimer AM, Wu SS, Gelfand AE, Silander JA, 2006. Building statistical models to analyze species distributions. *Ecological Application*, 16: 33-50.
- Liu C, White M, Newell G, 2013. Selecting thresholds for the prediction of species occurrence with presence-only data. *Journal of Biogeography*, 40: 778-89.
- Mackenzie DI, Nichols JD, Lachman GB, Droege S, Royle JA, Langtimm CA, 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83: 2248-2255.
- Massei G, Kindberg J, Licoppe A, Gačić D, Šprem N, Kamler J, Baubet E, Hohmann U, Monaco A, Ozoliņš J, Cellina S, Podgórski T, Fonseca C, Markov N, Pokorny B, Rosell C, Náhlik A, 2015. Wild boar populations up, numbers of hunters down? A review of trends and implications for Europe. *Pest Management Science*, 71: 492-500.
- Mammal Diversity Database (MDD), 2020. Mammal Diversity Database (Version 1.2) [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.4139818>. Map of Life. 2021. Mammal range maps harmonised to the Mammals Diversity Database [Data set]. Map of Life.
- Neftalí S, Barbosa AM, 2021. Common mistakes in ecological niche models, *International Journal of Geographical Information Science*, 35: 213-226.

- Pearce J, Ferrier S, 2001. The practical value of modelling relative abundance of species for regional conservation planning: a case study. *Biological Conservation*, 98: 33-43.
- Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S, 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, 19: 181-197.
- Pittiglio C, Khomenko S, Beltran-Alcrudo D, 2018. Wild boar mapping using population-density statistics: From polygons to high resolution raster maps. *PloS One*, 13: e0193295.
- R Development Core Team, 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.
- Ranc N, Santini L, Rondinini C, Boitani L, Poitevin F, Angerbjörn A, Maiorano L, 2017. Performance trade-offs in target-group bias correction for species distribution models. *Ecography*, 40: 1076-1087.
- Royle JA, 2006. Site Occupancy Models with Heterogeneous Detection Probabilities. *Biometrics*, 62: 97-102.
- Vieilledent G, Merow C, Guélat J, Latimer AM, Kéry M, Gelfand AE, Wilson AM, Mortier F, Silander Jr, JA, 2014. hSDM: hierarchical Bayesian species distribution models. R package version 1.4. <https://CRAN.R-project.org/package=hSDM>.
- Wilson DE, Lacher Jr TE, Mittermeier RA, Rylands AB (Eds.), 2009-2019. Handbook of the Mammals of the World: (Vol. 1 - Vol. 9). Barcelona: Lynx Editions. Map of Life. 2021. Mammal range maps digitized from the Handbook of the Mammals of the World (Wilson et al., 2009-19) [Data set]. Map of Life.
- Zuur AF, Ieno EN, Elphick CS, 2010. A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1: 3-14.

Glossary

Absolute population density	Number of individuals per surface unit, usually by km ² . This is an absolute measure that allows to make direct comparison among populations.
Abundance estimate	The number of individuals in a population calculated by statistical methods.
ASF	African Swine Fever.
AUC	Area Under Curve. Refers to the area under a Receiver Operator Curve (ROC) plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The resulting value reflects the predictive accuracy of a model where 0.5 indicates predictions no better than random (i.e., uninformative) and 1 indicates perfect prediction. Typically, values of 0.7 or greater are considered an indication of good performance.
Bioregion	Homogeneous bioclimatic regions based on bioclimatic variables, vegetation cover and topographic covariates associated to wild boar density
Cross-validation	Method of evaluating predictive models by partitioning sample data into a training set to fit the model and a testing set to use for evaluation.
Downscaling	It is a procedure to obtain predictions from a statistical model at a higher spatial resolution than used to parameterize the model. In this case, models were transferred from NUTS3 to UTM 10x10 km resolution.
Environmental domain	The range of environmental predictors that is included in the training datasets. That is, if you train a model within a range 2-20°C of temperature, the model only can explain the response to the species to that range, but the model does not have information about how the species is able to respond in localities without that range.
GBIF	Global Biodiversity Information Facility.
GLM	Generalised Linear Model.
Habitat suitability	The ability of a habitat to provide a species life requisites under current conditions.
Hunting bag	It refers to the number of animals hunted in a territory usually during a given hunting season.
Hunting yield	It is usually used to refer to a relative abundance index based on hunting bag data.
IUCN	International Union for Conservation of Nature.
MESS	Multivariate Environmental Similarity Surface.
Model extent	This term refers to the geographical area on which the model is to be fitted.
NUTS3	Nomenclature of Territorial Units for Statistics Level 3.
Population density (d)	It is a measurement of population size per area unit, i.e., population size divided by total land area. The absolute density usually is expressed in heads per 100 ha. Multiplying the population density by

	the studied surface, we obtain the population size. It can be calculated by different methods (either direct or indirect, summarized in Table 1).
Population size or absolute abundance (N)	It is the size of the population. It can be a known or estimated number, expressed in number of individuals. When related to area unit it gives the population density.
Predictive accuracy	Quantitative metric describing the accuracy of model predictions. Computed by comparing model predictions against independent data often obtained through a process of cross-validation.
Presence-absence	Dataset contain independent locations with binary classification describing whether a species is present or explicitly absent.
Presence-background	Independent datasets describing environmental conditions at locations where a species has been observed and those of a random sample from the available landscape.
Presence-only	Dataset containing independent events (date, location, recorder) describing species sightings; positive occurrences.
Relative abundance	Index describing the difference in populations across locations. Typically expressed using a discrete classification scale. When expressed as a continuous scale relative abundance can be transformed in absolute abundance using a population count at a single location.
Relative score	Index describing the difference in suitability, i.e., likelihood of species presence, across locations.
RSF	Resource Selection Function.
Suitability	Measure of how suitable a location is for a particular species; analogous to the likelihood that a species is present.
Training dataset	Split the dataset is a common modelling practice aimed to use a proportion of data to fit the model (training dataset) and the rest of data to assess the model performance on independent (i.e., not use in model fitting) data (evaluation dataset).
TSS	True Skill Statistic is a measure of model accuracy which considers omission and commission errors, and success because of random guessing. It ranges from -1 to $+1$, ($+1$ indicates perfect agreement and values of zero or less indicate a performance no better than random). TSS is not affected by prevalence neither by size of the validation set.
Variable importance	Quantitative measure of the relative importance/contribution of model variables in explaining observed data.