



# EXTERNAL SCIENTIFIC REPORT

APPROVED: 3 October 2022

doi:10.2903/sp.efsa.2022.EN-7631

## New models for wild ungulates occurrence and hunting yield abundance at European scale

ENETWILD-consortium<sup>1</sup>, Sonia Illanas, Simon Croft, Graham C. Smith, Sergio López-Padilla, Joaquín Vicente, José Antonio Blanco-Aguar, Massimo Scandura, Marco Apollonio, Ezio Ferroglio, Stefania Zanet, Rachele Vada, Oliver Keuling, Kamila Plis, Tomasz Podgorski, Francesca Brivio, Javier Fernández-López, Carmen Ruiz-Rodríguez, Ramón C. Soriguer, Pelayo Acevedo

### Abstract

The goal of this report is i) to model the occurrence and hunting yield (HY) density of wild ungulates not only for widely distributed species in Europe, but also for those ones which have a constrained distribution and ii) to compare the output of occurrence with observed HY. Random Forest function was used for modelling occurrence of species. We used occurrence data available from the past 30 years, and HY data (period 2015-2020) from records collected by *ENETWILD*. Like previous models based on HY, the response variable was the maximum number of wild ruminants annually hunted in 2015-2020 hunting seasons divided by the area (km<sup>2</sup>) of the corresponding administrative unit (HY density). Models based on HY were statistically downscaled to make predictions to 10x10km squares. Occurrence data models indicated a good predictive performance for most species, showing that the model framework proposed have improved results in comparison to previous models. The transferability of models into new regions was limited by the exposure of species to environmental conditions. As for HY models, the calibration plots showed a good and linear predictive performance for widely distributed species, as well as constrained distributed species. Overall, our results were consistent with the expected abundance distribution of widely distributed species. The removal of zeros on the validation datasets affected the calibration plots of all regions, showing a better predictive performance when zeros were removed for widely distribution species, but the opposite was evidenced for species with limited distributions. We conclude that (i) the importance of co-correlation variables when variable importance is inferenced from random forest model results, (ii) manipulation presence and absence locations could yield further improvement in occurrence model outputs, and (iii) HY model projections displayed good abundance patterns for most of species, showing that the three frameworks proposed were a good approximation for modelling the distribution of wild ungulates HY, although it should be explored how to improve the results when distribution is patchy.

© European Food Safety Authority, 2022

**Key words:** distribution, hunting bags, occurrence, population abundance, population monitoring, spatial modelling, wild ungulates

**Question number:** EFSA-Q-2022-00045

**Correspondence:** [biohaw@efsa.europa.eu](mailto:biohaw@efsa.europa.eu)

<sup>1</sup> ENETWILD Consortium: [www.enetwild.com](http://www.enetwild.com)

**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 consortium, Sonia Illanas, Simon Croft, Graham C. Smith, Sergio López-Padilla, Joaquín Vicente, José Antonio Blanco-Aguiar, Massimo Scandura, Marco Apollonio, Ezio Ferroglio, Stefania Zanet, Rachele Vada, Oliver Keuling, Kamila Plis, Tomasz Podgorski, Francesca Brivio, Javier Fernández-López, Carmen Ruiz, Ramón C. Soriguer, Pelayo Acevedo. 2022. New models for wild ungulates occurrence and hunting yield abundance at European scale. EFSA supporting publication 2022:EN-7631. 43 pp. doi: 10.2903/sp.efsa.2022.EN-7631

**ISSN:** 2397-8325

© European Food Safety Authority, 2022

## Summary

**Background and objectives:** In *ENETWILD* consortium et al. (2021) the generic model framework for predicting 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) was defined. Across 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 and showed a similarly predictive accuracy, except for roe deer which showed a lower accuracy predicting absences. On the other hand, during the last year the consortium *ENETWILD* has continued collecting data on hunting yield (HY) for wild ruminants and wild boar over Europe at different spatial resolution (e.g., hunting ground, municipalities, NUT3), including species with a limited distribution. The goal of this report is modelling the occurrence and HY density of wild ungulates not only for widely distributed species in Europe, but also for those ones which have a constrained distribution, evaluating the performance of the approaches, and compare outputs (HY vs. occurrence), after the downscaling of HY models to make predictions to 10x10km squares.

**Data:** We used occurrences available from the past 30 years (1990- 2020) through the Global Biodiversity Information Facility (GBIF) extracted on 26/04/2022<sup>2</sup> together with records from *iMammalia* (*MammalNet* project <https://mammalnet.com/>; ENETWILD consortium et al., 2018), and *MammalWeb* (<https://www.mammalweb.org/en/>), and HY data compiled for the period 2015-2020 from records submitted to *ENETWILD* Data Model, extracted on 30/05/2022.

**Modelling:** Random Forest function was used for modelling occurrence of species. Like previous models based on HY, the response variable was the maximum number of wild ruminants annually hunted in 2015-2020 hunting seasons divided by the area (km<sup>2</sup>) of the corresponding administrative unit (HY density). We conducted negative binomial generalized linear models (GLMNB) for widely distributed species; concretely we developed four models (one per European bioregion) for wild boar and one model for all Europe for red deer, roe deer and fallow deer. Zero inflated negative binomial generalized ZINB models were conducted for species with a limited distribution. GLMNB and ZINB models included eco-geographical variables as predictors. 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:** The model performance of occurrence data models indicate a good predictive performance of all models exceeding the AUC value 0.7 for all species except one (*Ovis moschatus*, an introduced species with local distribution), showing that the current model framework has improved results in comparison to previous models. Other statistic metrics for model evaluation also showed an improvement, which was aligned with the visual inspection of predicted distributions. Model projections showed no signs of spatial autocorrelation in residuals. The transferability of models into new regions was limited by the exposure of species to environmental conditions, so species with small ranges have restricted transferability. As for HY models, the calibration plots showed monotonic relationships and a good predictive performance for widely distributed species, as well as for constrained distributed species. Although, it seemed that predictions at high densities could be improved for red deer, that there were some overprediction for roe deer and fallow deer, and that the predictive performance of chamois, reindeer, and sika deer models neither show a clear pattern nor a precise prediction. Moreover, the removal of zeros on the validation datasets affected the calibration plots of all regions, showing a better predictive performance when zeros were removed. Contrary to previously, calibration plots for constrained species showed the importance of maintaining zero values in the datasets for a good model prediction. Overall, our results were consistent with the expected

<sup>2</sup> <https://doi.org/10.15468/dl.g3jbgh>

abundance distribution of widely and constrained distributed species, except for mouflon and sika deer, which did not show a reliable pattern of abundance distribution. All species showed a different spatial pattern as well as different values of densities, being widely distributed species those ones with higher densities in Europe.

### Conclusions and next steps:

#### *Occurrence data model*

- A two-step occurrence model, identifying likely absence, and then exploring powerful machine learning algorithms to fit presence-absence data, yields improved predictions across all species compared with the previous coupled Bayesian framework. There are several plausible explanations for this, but one factor may be the ability of random forest to better capture non-linear relationships between occurrence (suitability) and environmental variables.
- The new approach allows the use of raw variables (rather than requiring an opaque PCA transformation to facilitate fitting; ENETWILD consortium et al., 2021). While random forest is somewhat of a “black-box” inference on variable importance, which could provide valuable insight/verification. Such outputs can however be influenced by factors like co-correlation (Strob et al., 2008). Further work is required to understand these issues before robust inference can be gained but potential solutions are available (party package in R; Strob et al., 2008).
- Identification of absence prior to modelling allows examination of the spatial structure and balance between presences and absences. It has been suggested that these patterns can impact model outputs, for instance where mixing is poor. Additional investigation is required to determine if manipulation (careful subsampling) of presence and absence locations could yield further improvement in model outputs (Steen et al., 2021).

#### *Hunting yield density data model*

- Model projections showed good abundance patterns for most of species. There is not a unique framework for modelling at European scale and it should be adapted to the specifics of the distribution of the dataset for modelling. The three frameworks proposed were a good approximation for modelling the abundance distribution of species.
- The zeros, i.e., the areas where the species is not hunted, are not informative for modelling species with wide distribution range and therefore should be removed from the dataset but they should be maintained when modelling species with constrained ranges.
- A patchy distribution of collected HY lead to a weird spatial pattern of abundance in several species (e.g. sika deer). For some species, better quality of data provided is still needed (amount and better spatial resolution if available) for modelling purposes (e.g., reindeer).

#### *Validation of suitability on HY*

- The comparison of observed HY and suitability predicted values are in consonance with hunting yields values for most of species. This may indicate that suitability has potential to be used as a proxy for abundance in this species. However, the resolution of intervals for using suitability as a proxy of abundance will depend on each species.
- The different relationships saw on species may suggest that incorporating habitat suitability values in the HY models as explanatory variable could not always increase the explained variance of the model and that it is needed to clarify that relations and determine if its inclusion could benefit HY modelling results in all species.

## Table of contents

Abstract .....	1
Summary .....	3
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 .....	11
2.2.1. Wild ruminant occurrence data .....	11
2.2.2. Hunting yield data .....	11
3. Methodology .....	15
3.1. Models based on occurrence data .....	15
3.2. Models based on hunting yield 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 .....	16
4.1. Occurrence data models .....	16
4.2. Models based on hunting yield data .....	21
4.3. Comparison of models based on hunting yield and occurrence data .....	35
4.4. Conclusions and further steps .....	38
References .....	39
Glossary .....	42

## 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 9) refer to deliverable 4.1: to improve the model for wild ruminant distribution/abundance, which is due on June 2022.

### 1.2. Scope of the report

The ENETWILD consortium ([www.enetwild.com](http://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 February 2022 (<https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/sp.efsa.2022.EN-7174>) ENETWILD enhanced the proposed 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 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, and showed a similarly predictive accuracy, except for roe deer, which showed a lower accuracy predicting absences. On the other hand, during the last year the consortium ENETWILD has continued collecting data on hunting yield (HY) for wild ruminants over Europe at different spatial resolution (e.g., hunting ground, municipalities, NUT3), including species with a constrained distribution. The goal of this report is modelling occurrence and HY density of wild ruminants, not only for widely distributed species in Europe, but also for those having a constrained distribution, evaluating the performance of the approaches, and comparing the outputs (HY vs occurrence), after the downscaling of HY models to make predictions to 10x10km squares.

### 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 3.22.9, tidyverse (Wickham et al., 2019) and *sf* (Pebesma, 2018) R packages.

**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.

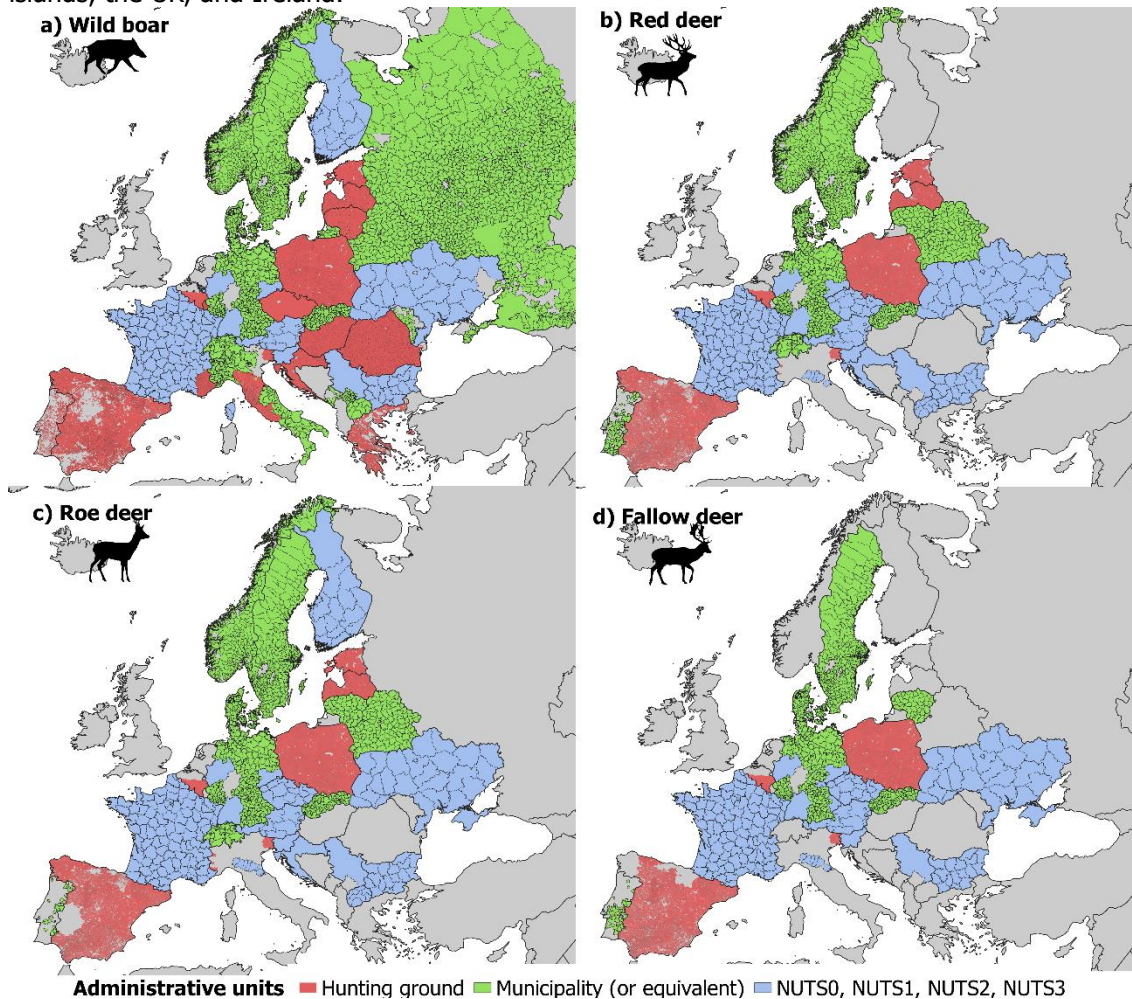
<b>Code</b>	<b>Variable description</b>	<b>Code</b>	<b>Variable description</b>
BIO1	Annual mean temperature	lc_10	Cropland, rainfed
BIO2	Mean diurnal range (mean of monthly (max temp - min temp))	lc_11	Herbaceous cover
BIO3	Isothermality (BIO2/BIO7) (x 100)	lc_12	Tree or shrub cover
BIO4	Temperature 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 sp.</i>	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

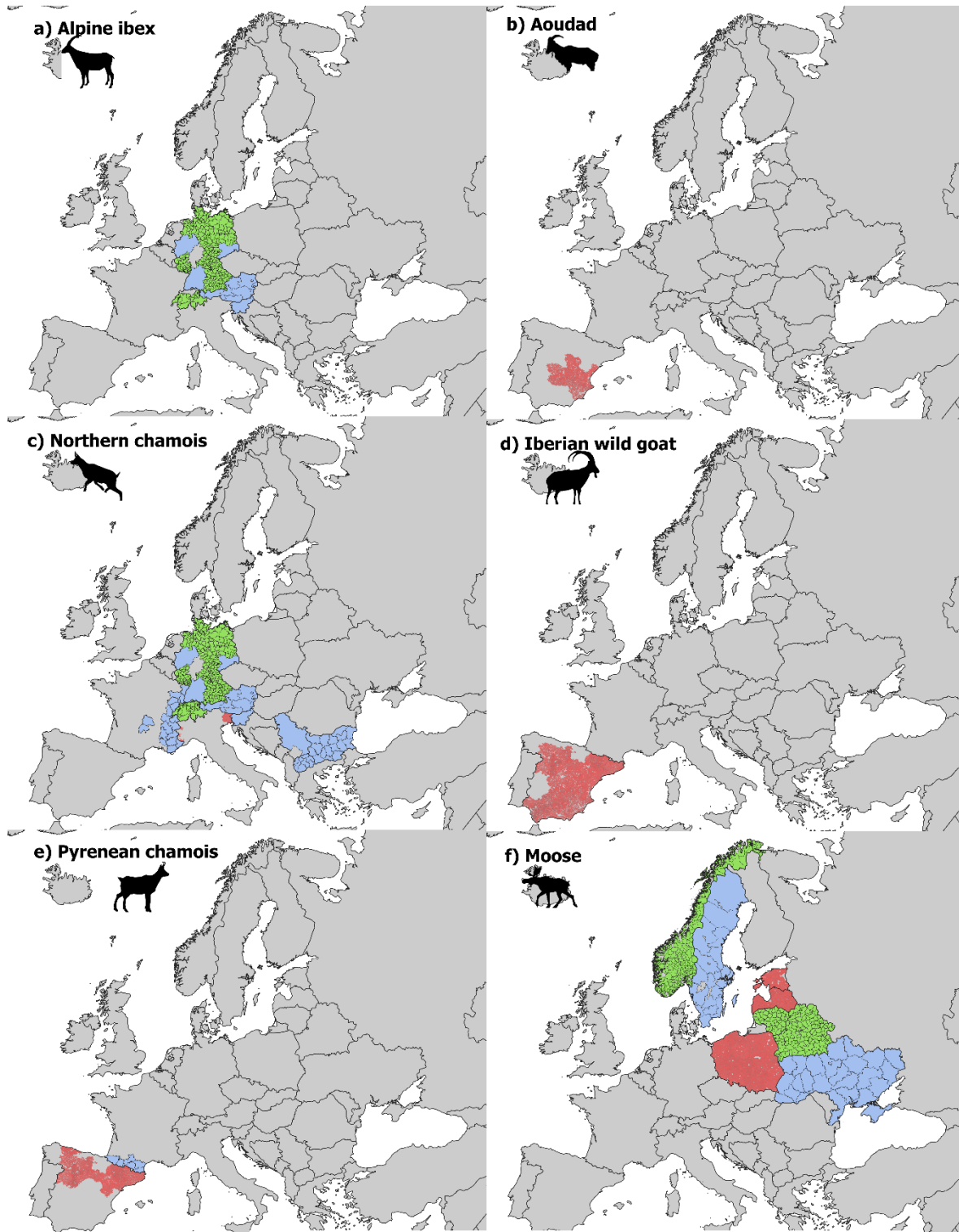
### 2.1. Study area

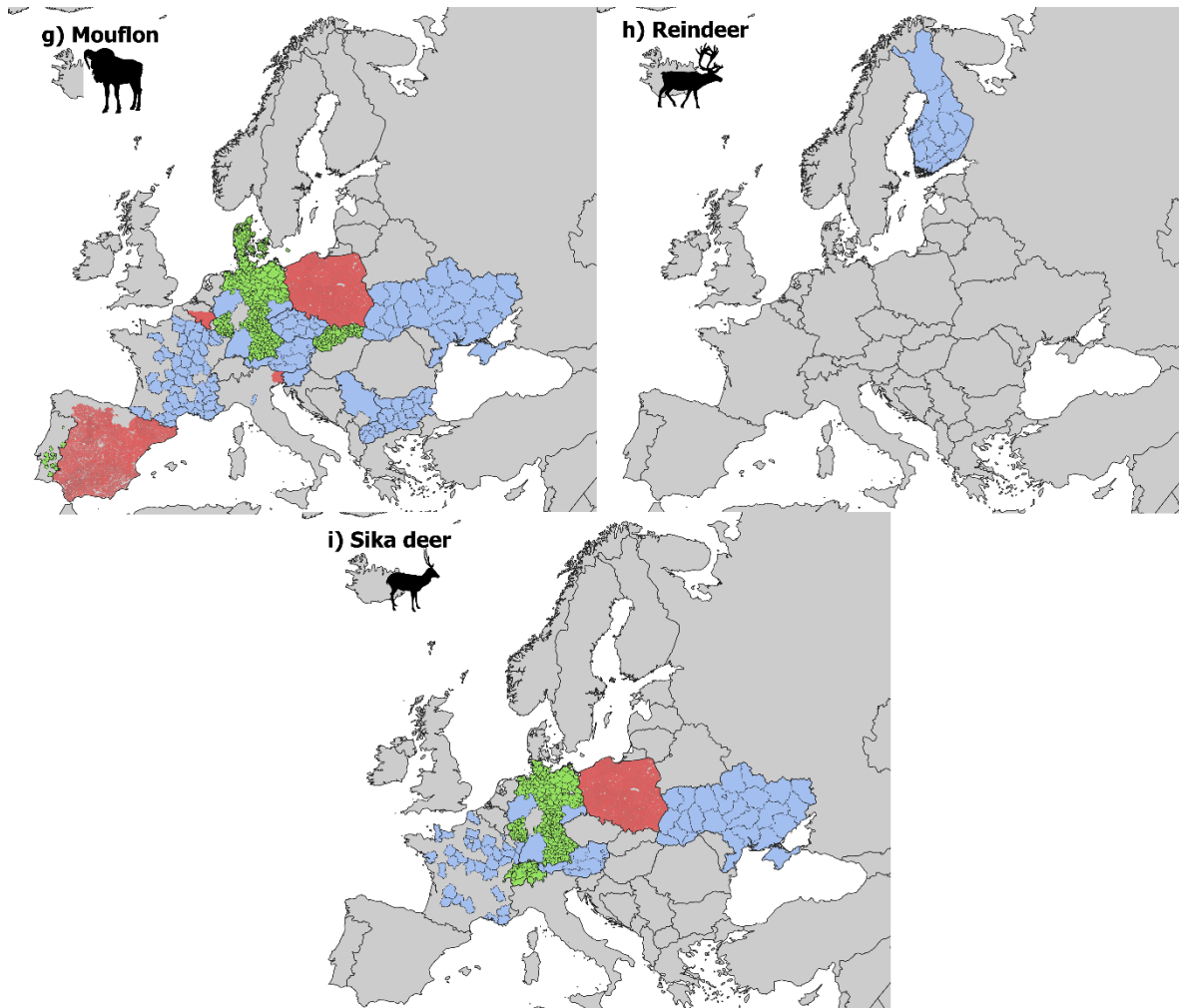
The study area is the same as the previous report (ENETWILD consortium et al., 2021). It includes all countries in mainland Europe with the Ural Mountains as the eastern limit (Figure 1), spans 11,019,700 km<sup>2</sup> (110,197 10x10 km and 2,787,877 2x2 km grid cells) and includes Mediterranean islands,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



**Figure 1:** Maps 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 widespread species according to hunting data availability (territorial units, see below section 2.2.2): (a) wild boar, (b) red deer, (c) roe deer, and (d) fallow deer hunting yield.







**Administrative units** ■ Hunting ground ■ Municipality (or equivalent) ■ NUTS0, NUTS1, NUTS2, NUTS3

**Figure 2:** 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 constrained species according to hunting data availability (territorial units, see below section 2.2.2): (a) Alpine ibex, (b) Aoudad, (c) Northern chamois, (d) Iberian wild goat, (e) Pyrenean chamois, (f) moose, (fg) muntjac, (h) reindeer, and (i) sika deer hunting yield.

## 2.2. Data collection

### 2.2.1. Wild ruminant occurrence data

For modelling species occurrence, we obtained sightings records (presences) from the past 30 years (1990-2020) from collections hosted on the Global Biodiversity Information Facility (GBIF; extracted on 26/04/2022 <https://doi.org/10.15468/dl.g3jbgh>) and combined these with records from our own *MammalNet* data collection project (<https://mammalnet.com>, *iMammalia* app; ENETWILD consortium et al., 2018), and *MammalWeb*. Any records without an exact 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.

The resulting dataset was then processed in two ways to facilitate different functions of our modelling, creating: (i) a presence-only dataset to supplement expert drawn estimates of species range; (ii) a binomial dataset coupling positive sightings against an estimate of survey effort (visits defined with unique date and 2x2 km grid cell) derived by aggregating sightings across multiple "associated" species (see ENETWILD consortium et al., 2022 for more details). For the latter dataset we applied a more recent temporal range of 2011-2020 to provide better alignment with that of our explanatory variables (Table 3). Furthermore, to ensure any constituent collections included in this part of the analysis complied with the idea of multi-species recording we assessed each separately and excluded any which contained observations of fewer than 10 different species or 2 taxonomic orders.

### 2.2.2. Hunting yield data

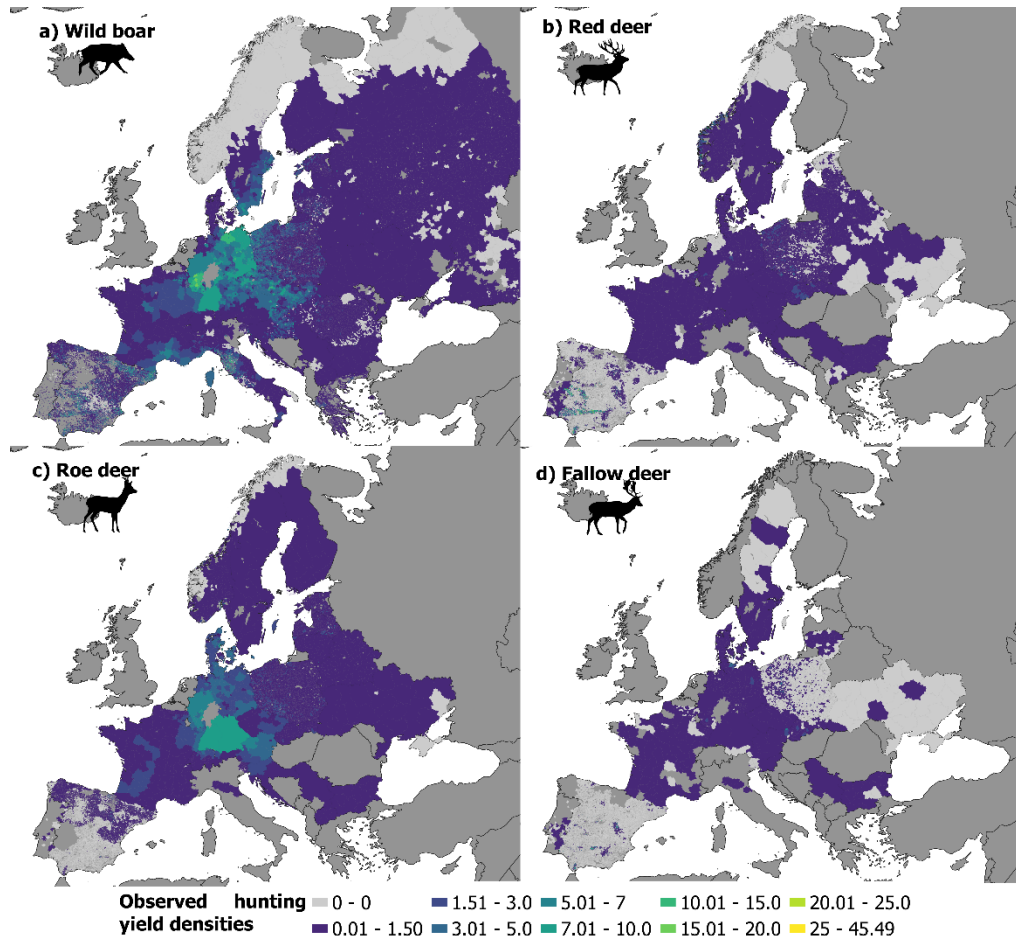
For this report hunting yield data were incorporated for modelling from the ENETWILD data collection. Their spatial resolution as well as number of records collected depend on the species (Figures 1 and 2, Table 2). We normally have more data, as expected, when species are widespread (e.g., wild boar, red deer, roe deer, fallow deer) than when they are more spatially constrained (e.g. Northern chamois, Pyrenean chamois, Iberian wild goat).

Like previous reports, we focused on maximum hunting yield records from 2015 to 2020 hunting seasons. Due to the later, 307 data reported by Sweden corresponding to 2019 could not have been considered as they correspond to municipality level while previous data of Sweden corresponded to lower spatial resolution (NUTS 3) and were therefore included. The response variable was obtained by dividing the maximum number of hunted animals by the area of the respective territorial unit (km<sup>2</sup>), i.e., we modelled hunting yield density (HY) of wild boar, red deer, roe deer, fallow deer, Alpine ibex, aoudad, Northern chamois, Iberian wild goat, Pyrenean chamois, moose, muntjac, reindeer, and sika deer (Figure 3 and 4).

Zero values remained for modelling for species with a constrained distribution only in the territories where hunting bags were reported as they are very informative, while zero values were removed for widespread species. Hunting yield density records were transformed to density data multiplying their values by 10000 for modelling purposes (to have integer response variable for the negative binomial models).

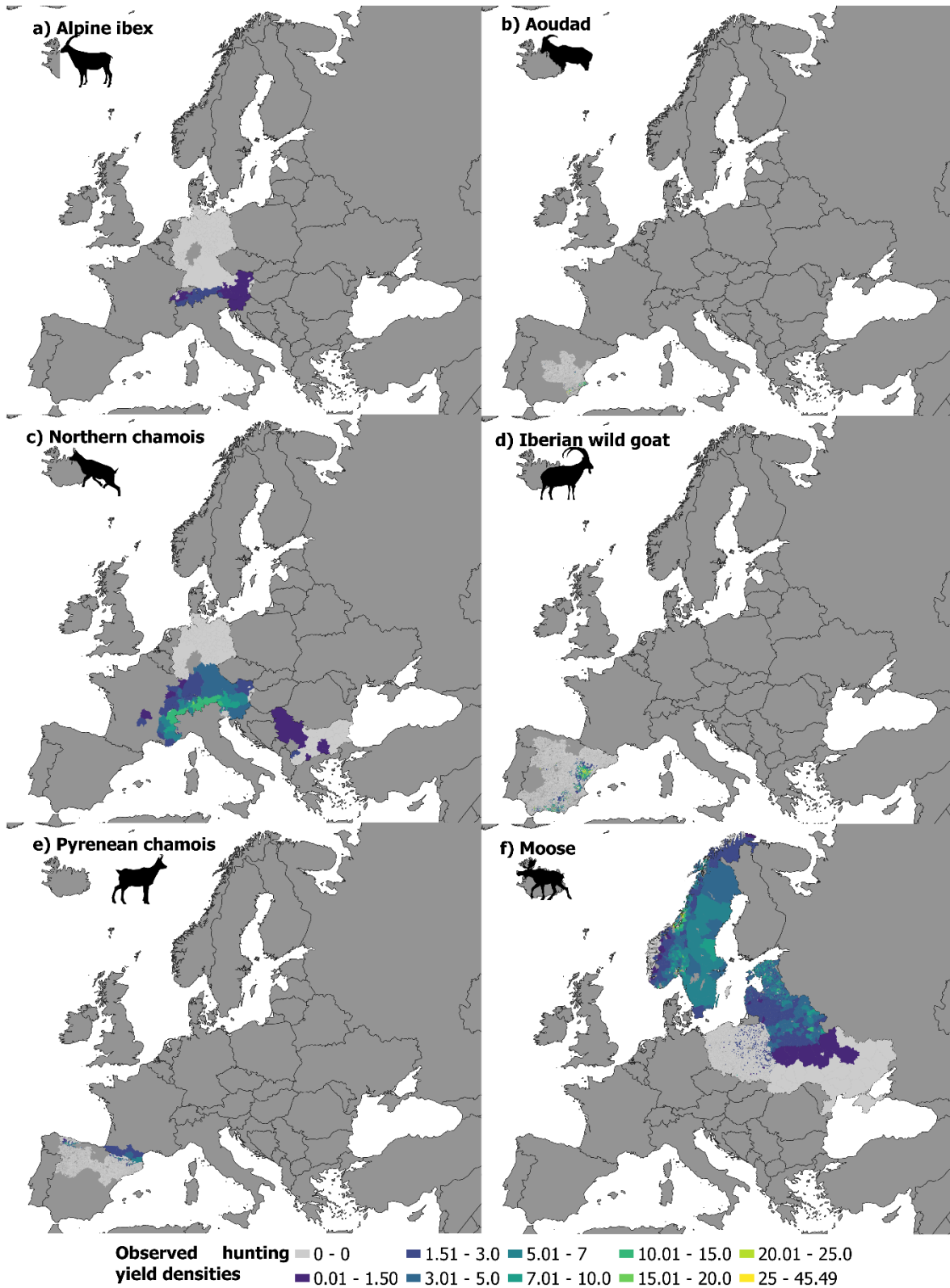
**Table 2.** Number of territorial units used for modelling by species.

	Country	Territorial units			Total	Percentage of 0 values
		Hunting ground	Municipality (or equivalent)	NUT0, NUT1, NUT2, NUT3		
Widespread species	Wild boar	32475	2184	290	34949	12.49%
	Red deer	33247	821	459	34527	69.88%
	Roe deer	30166	752	482	31396	58.50%
	Fallow deer	31313	269	419	32001	91.86%
Spatially constrained species	Alpine ibex	-	15	279	294	93.20%
	Aoudad	6647	-	-	6647	97.23%
	Northern chamois	321	26	331	678	75.81%
	Iberian wild goat	22170	-	-	22170	93.77%
	Pyrenean chamois	7256	-	6	7259	96.57%
	Moose	5758	525	48	6331	74.43%
	Mouflon	28435	192	380	29007	95.64%
	Reindeer	-	-	15	15	73.3%
Sika deer	4969	26	338	5333	98.44%	

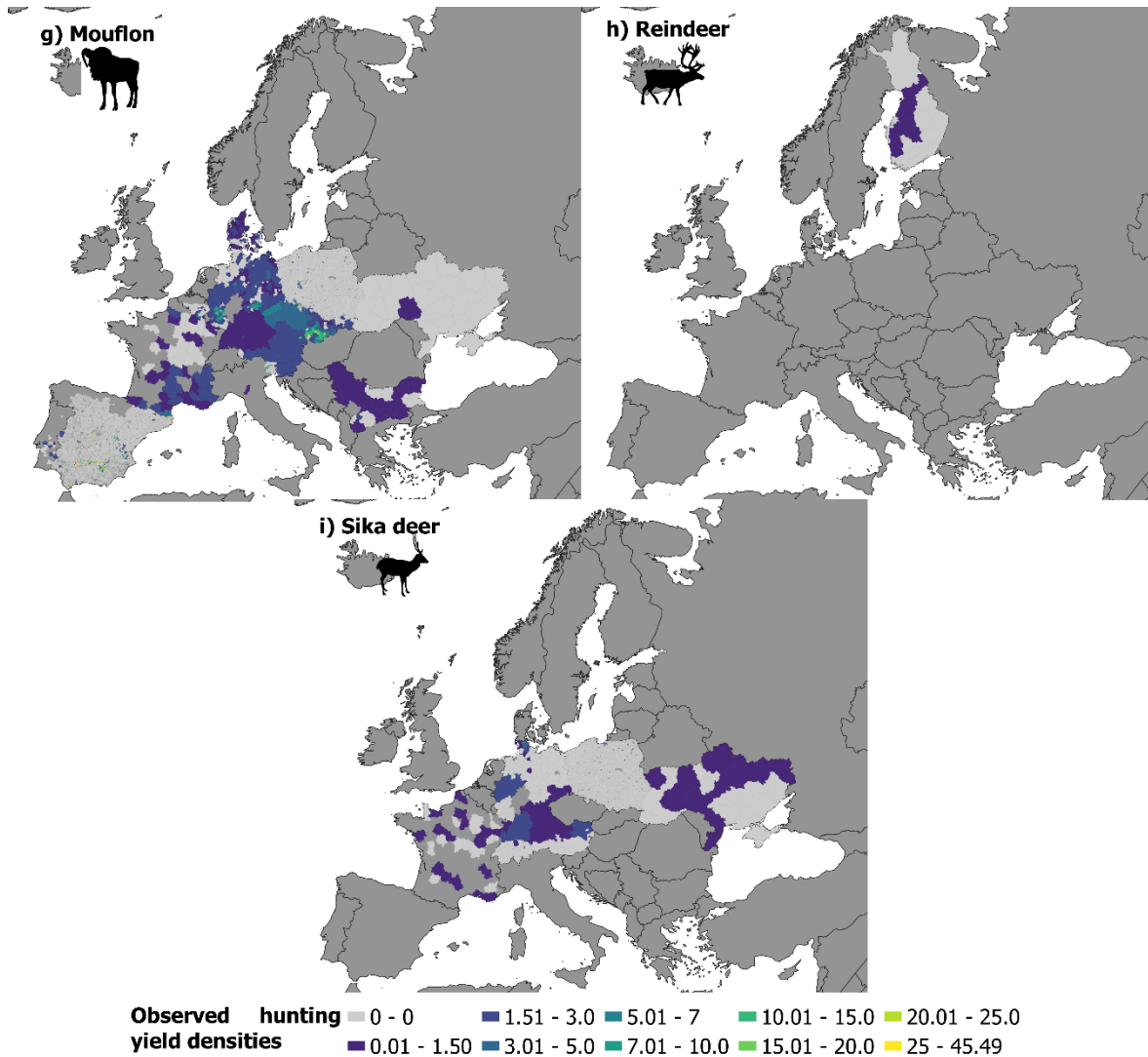


**Figure 3:** Maximum hunting yield density data (2015-2020 hunting seasons) used for parameterizing ungulates species models of widespread species according to territorial units (observed hunting bag density per km<sup>2</sup>).









**Figure 4:** Maximum hunting yield density data (2015-2020 hunting seasons) used for parameterizing ungulates species models of constrained species according to territorial units (observed hunting bag density per km<sup>2</sup>).

### 3. Methodology

#### 3.1. Models based on occurrence data

The model framework comprised two steps. Firstly, we estimated the broad-scale “stable” range for each species based on a combination of expert-derived maps (Burgin et al., 2020; MDD, 2020; IUCN 2021; Wilson et al., 2021) and maps derived from our presence-only dataset outlining regions with consistent sightings (Maes et al., 2015). Then, using the resulting ranges to filter our binomial dataset, we estimated a finer-scale (2x2 km resolution) description of occurrence (ensuring the condition of stationarity was observed i.e., we excluded influence from sites where populations may be in flux and therefore not wholly representative of the environmental conditions; Hattab et al., 2017).

In recent reports (ENETWILD consortium et al., 2021; 2022) we have advocated for an approach using a hierarchical Bayesian framework simultaneously fitting two processes, one describing the species ecology (occurrence) and one to account for observability (detection). However, following extensive testing, here we return to a more classical stepwise approach (as this improves the statistical measures of fit), estimating locations of likely absence based on survey effort (e.g., ENETWILD consortium et al., 2019) followed by presence-absence modelling using powerful machine learning techniques (Random Forest; Liaw & Wiener, 2002) which have previously been shown to perform well across a wide range of species (Croft et al., 2017). To fit these models of occurrence, we consider explanatory variables describing climate, land cover, topography, and human disturbance as in previous ENETWILD reports (Table 1). To mitigate co-correlation amongst variables which can impact model inference we identify and remove highly correlated variables based on variable inflation factor (VIF; values above 10 are indicative of problems with co-correlation) using the `usdm` package in R (Naimi et al., 2014).

For each species we use fitted models to project likelihood of occurrence (suitability) across the full extent of interest (Figure 5). We assess model performance using 4-fold cross validation, systematically reserving a proportion of each presence-absence dataset (25%) for testing, and applying the model fitting process described above to the remainder to predict occurrence. We compute several common metrics for predictive accuracy aggregated across folds (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). We also test model residuals for presence of spatial autocorrelation which can indicate missing explanatory variables and has the potential to confound evaluation and model inference using mantel correlograms (Legendre & Fortin, 1989). Finally, to evaluate the transferability of our model predictions we perform a MESS analysis (Elith et al., 2010) which identifies regions whose environmental conditions are deemed insufficiently represented by the training dataset so as not to produce unreliable prediction.

#### 3.2. Models based on hunting yield data

The response variables for modelling wild ungulates were hunting yield density across Europe, (maximum number of individuals annually hunted within 2015-2020 hunting seasons, divided by unit area in km<sup>2</sup>; hereafter HY).

We calibrated one model per species following two approaches depending on how wide or constrained the distribution of species was.

- For widely distributed species we followed the generic framework developed by ENETWILD consortium et al. (2021) for wild boar and by ENETWILD consortium et al. (2022) for red deer, roe deer, and fallow deer, eco-geographical predictors more relevant in explaining HY were determined using a generalized linear model (negative binomial distribution and logarithmic link function; Cameron & Trivedi 2013).

- For constrained distributed species (*i.e.*, Alpine ibex, aoudad, Northern chamois, Iberian wild goat, Pyrenean chamois, moose, mouflon, reindeer, and sika deer) we developed a new generic framework, in which zero values of the response variable remained in territories where hunting bags have been reported, instead of being removed as for widespread species. The same eco-geographical predictors which relevance are determined by a Zero Inflated Negative Binomial model were used (Ridout, Hinde and Demétrio, 2001).

Following steps were the same as in previous reports. 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 & Ferrier, 2001), where monotonicity of the relationship informs about the reliability of the predicted pattern. Moreover, for widespread species, 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>).

### 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 predicted 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

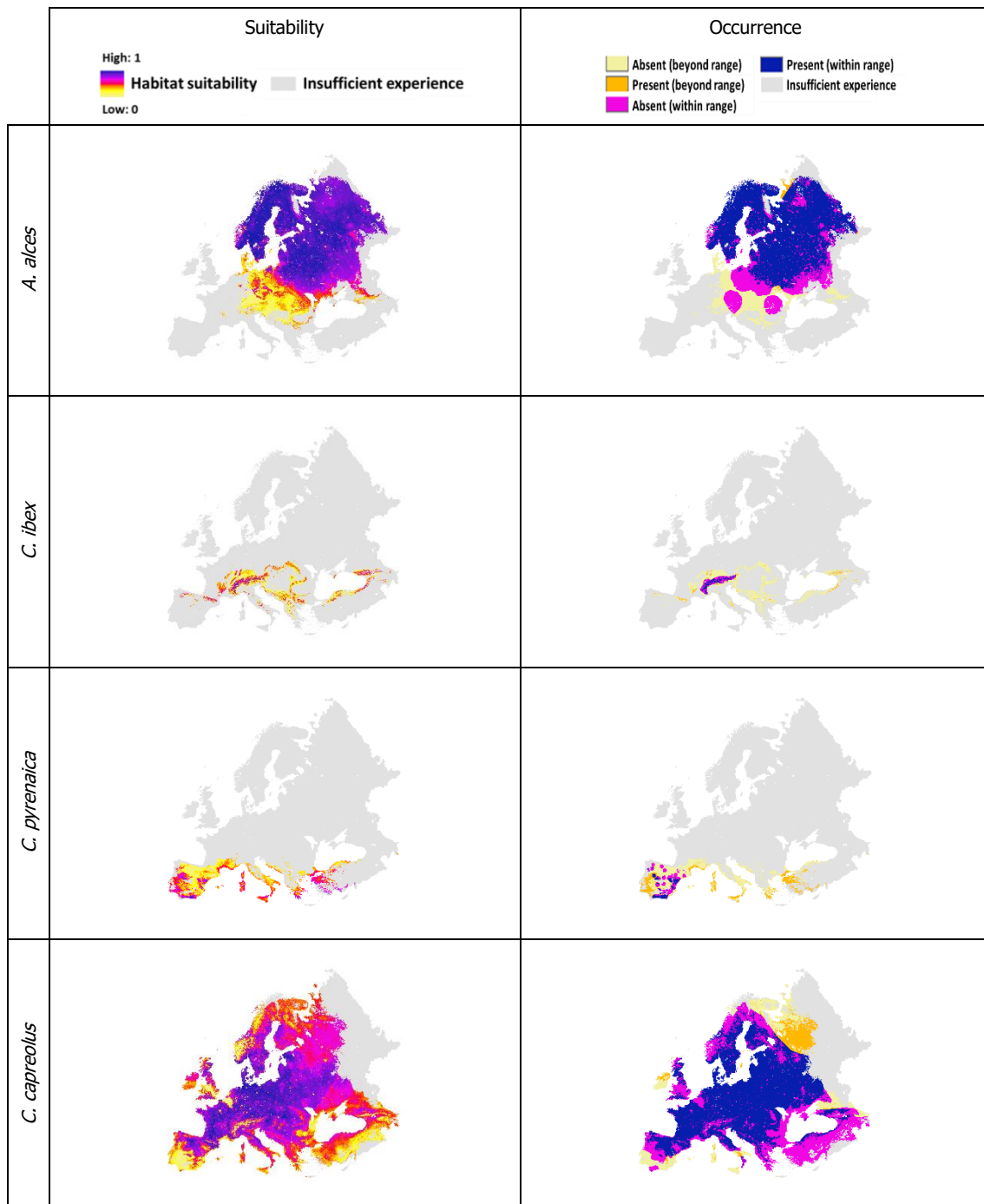
### 4.1. Occurrence data models

Sufficient data were available to make predictions for 15 of the 19 *Artiodactyla* species present in Europe (ruminants plus wild boar) considered by this study. Of those where prediction (and evaluation) was possible our results showed a marked improvement in evaluation statistics compared with previous reports (Table 3). The AUC statistic for all species except for *O. moschatus* exceeded 0.7 indicating good predictive performance reaching as high as 0.89 for *O. virginianus*. Other threshold-dependent metrics (sensitivity, specificity, and True Skill Statistic) were also increased with greater balance between sensitivity and specificity (*i.e.*, equally good prediction of absence as presence). The improvement in evaluation metrics aligned with the visual inspection of predicted distributions (Figure 5). Comparison against existing projections (Croft et al., 2019), albeit at a local extent (UK), showed good agreement in spatial patterns across species. Model projections showed no signs of spatial autocorrelation in residuals which given the variation in environmental conditions is perhaps unsurprising. As we have seen previously the

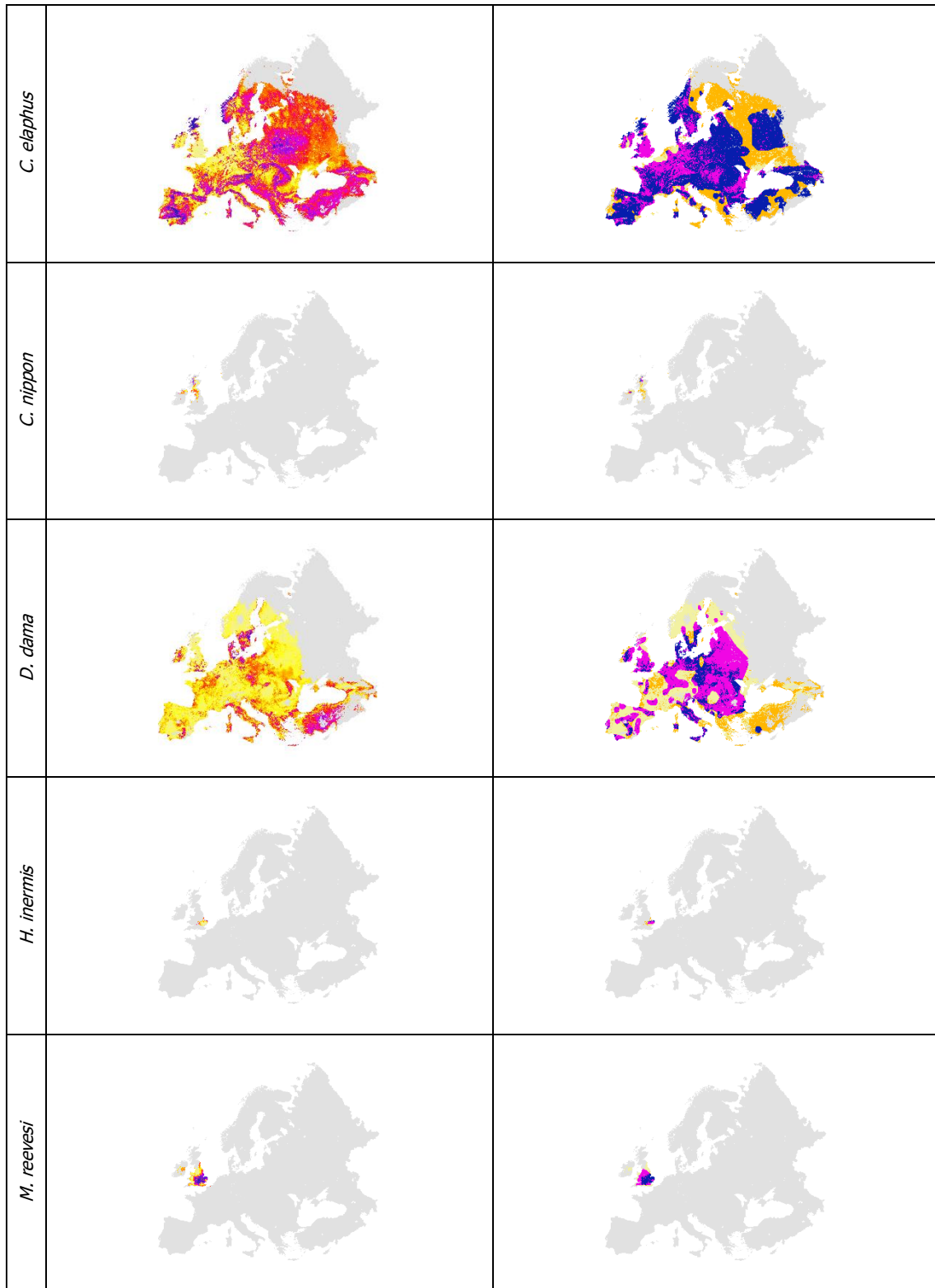
transferability of models, i.e., the reliability of projections into new regions, is limited by the exposure of species to environmental conditions. As such, projections for localised species with small ranges are much more restricted transferability (Figure 5).

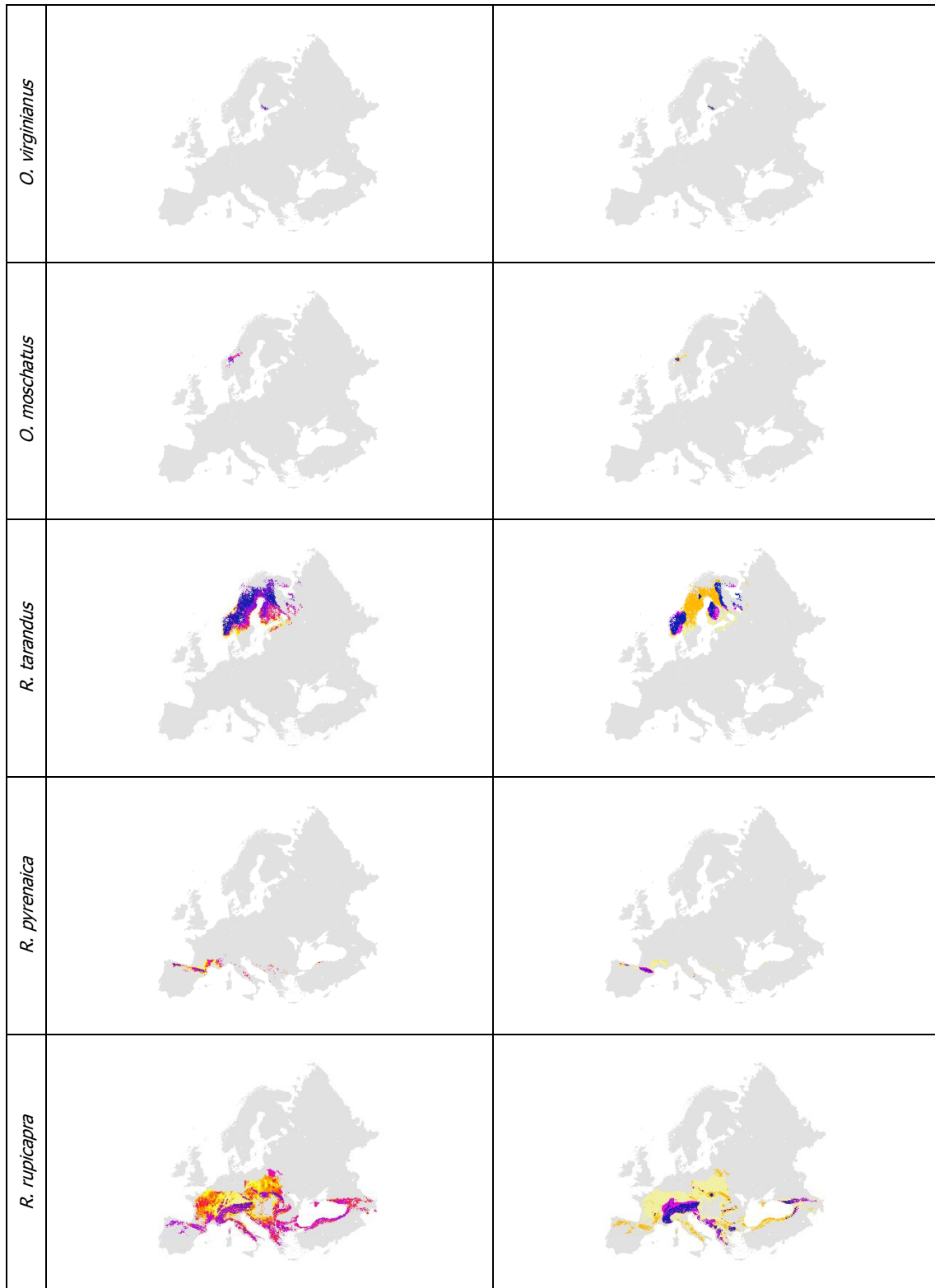
**Table 3.** Summary of distribution statistics (estimated “stable” range size, occupancy within range and potential occupancy across the model extent derived from thresholded model outputs, Figure 5, occurrence, including areas where prediction may be uncertain, in km<sup>2</sup>), predicted observability (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).

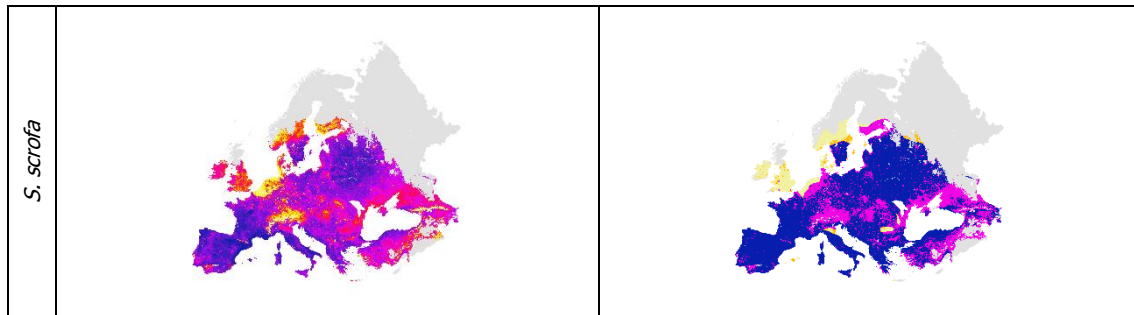
Species	Rng	Occ.	Pot.	Obs.	AUC	THD	SE	SP	TSS
<i>A. alces</i>	1,158,5680	3,943,368	3,990,712	0.15	0.87	0.80	0.82	0.78	0.60
<i>A. lervia</i>	-	-	-	-	-	-	-	-	-
<i>A. axis</i>	-	-	-	-	-	-	-	-	-
<i>B. bonasus</i>	-	-	-	-	-	-	-	-	-
<i>C. aegagrus</i>	-	-	-	-	-	-	-	-	-
<i>C. ibex</i>	212,544	42,368	2,925,528	0.37	0.74	0.52	0.76	0.60	0.36
<i>C. pyrenaica</i>	380,824	63,488	1,693,224	0.63	0.77	0.47	0.76	0.72	0.48
<i>C. capreolus</i>	14,336,928	4,584,480	5,751,216	0.37	0.82	0.59	0.85	0.65	0.50
<i>C. elaphus</i>	11,320,936	4,103,020	8,293,892	0.21	0.86	0.34	0.81	0.76	0.57
<i>C. nippon</i>	20,696	6,124	9,078,384	0.21	0.78	0.33	0.79	0.69	0.48
<i>D. dama</i>	5,861,256	811,032	3,051,068	0.24	0.81	0.34	0.69	0.79	0.48
<i>H. inermis</i>	23,176	3,444	8,889,356	0.18	0.76	0.29	0.81	0.69	0.51
<i>M. reevesi</i>	196,008	45,852	61,060	0.14	0.79	0.60	0.78	0.69	0.47
<i>O. virginianus</i>	16,808	7,768	6,733,508	0.39	0.89	0.49	1.00	0.58	0.58
<i>O. moschatus</i>	9,856	3,916	328,160	0.86	0.67	0.73	0.82	0.51	0.33
<i>R. tarandus</i>	2,095,776	850,812	1,745,896	0.04	0.85	0.69	0.93	0.80	0.73
<i>R. pyrenaica</i>	68,992	10,448	66,128	0.42	0.76	0.78	0.65	0.78	0.43
<i>R. rupicapra</i>	947,496	204,712	4,050,644	0.45	0.78	0.60	0.80	0.66	0.47
<i>S. scrofa</i>	13,982,472	437,9320	4,893,672	0.14	0.86	0.65	0.78	0.77	0.56











**Figure 5:** Predictions of wild ruminant and wild boar 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. The current range of species distribution can be seen in Fig. 8.

#### 4.2. Models based on hunting yield data

The model outputs are available at this [link](#).

The calibration plots of wild boar (Figure 7a) showed a good predictive performance for all bioregions.

Red deer HY density model (Figure 7b) showed a good predictive performance for almost all bins, although latest bins showed overfitted or underfitted patterns, which may indicate that high densities predictions could be improved. 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 and Northern bioregion showed a good predictive performance, and the predictive performance of Northern bioregion in relation to previous report were improved. However Western bioregion still showed a bad predictive performance.

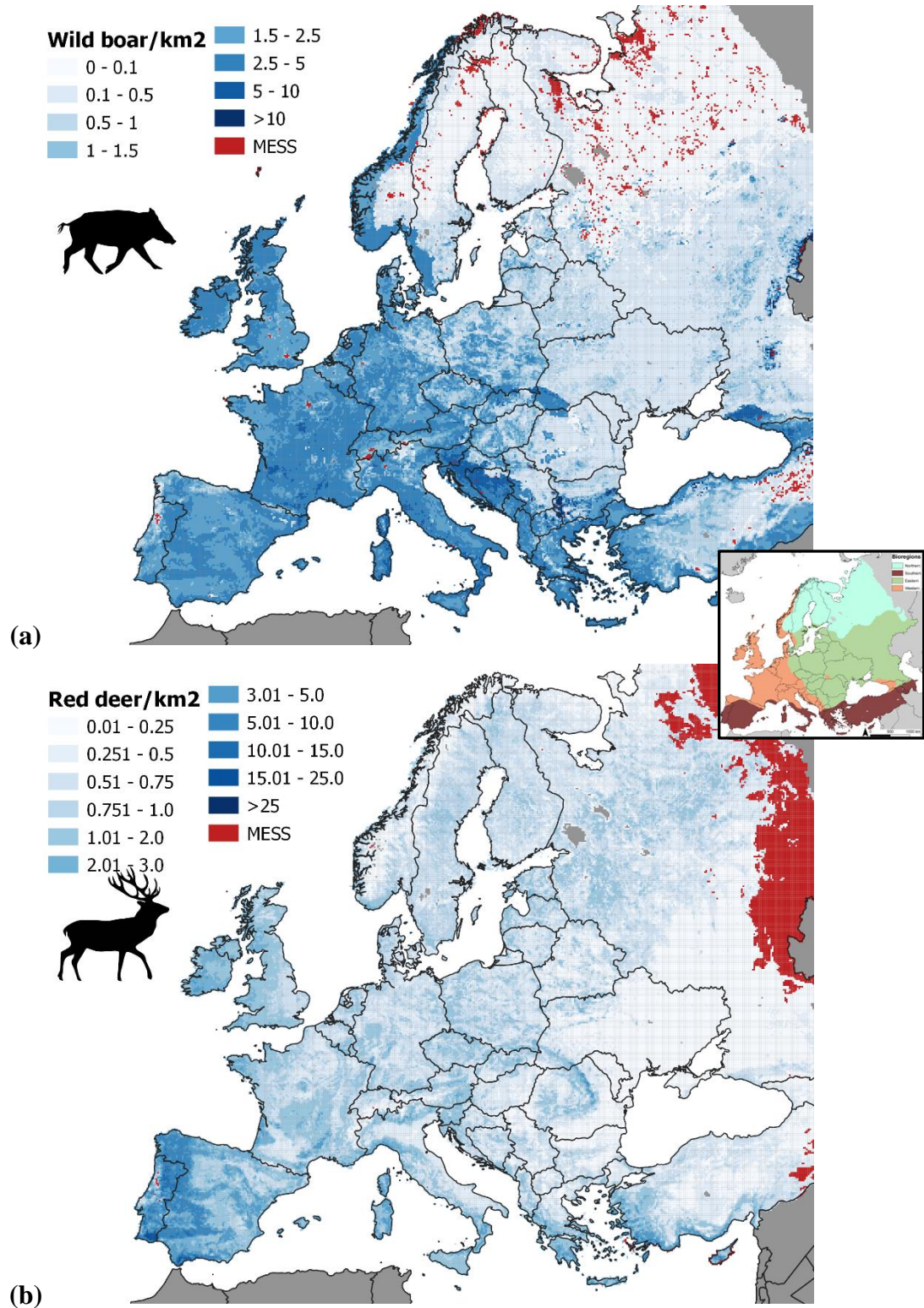
The general calibration plot for roe deer HY model (Figure 7c) showed an overpredicted pattern, and it mostly had a pattern like that of the Southern bioregion, like the red deer. An indetermined pattern is found in Northern bioregion, showing under or overprediction depending on the bin. Better predictive performance is found at Eastern and Western bioregions, showing both a good predictive performance. Moreover, predictive performance seemed to have improved in comparison to the previous model.

The general calibration plot for fallow deer HY model (Figure 7d) showed an overpredicted pattern. The pattern was mostly like the Southern and Western bioregions, and we only found a good predictive performance for Eastern bioregion.

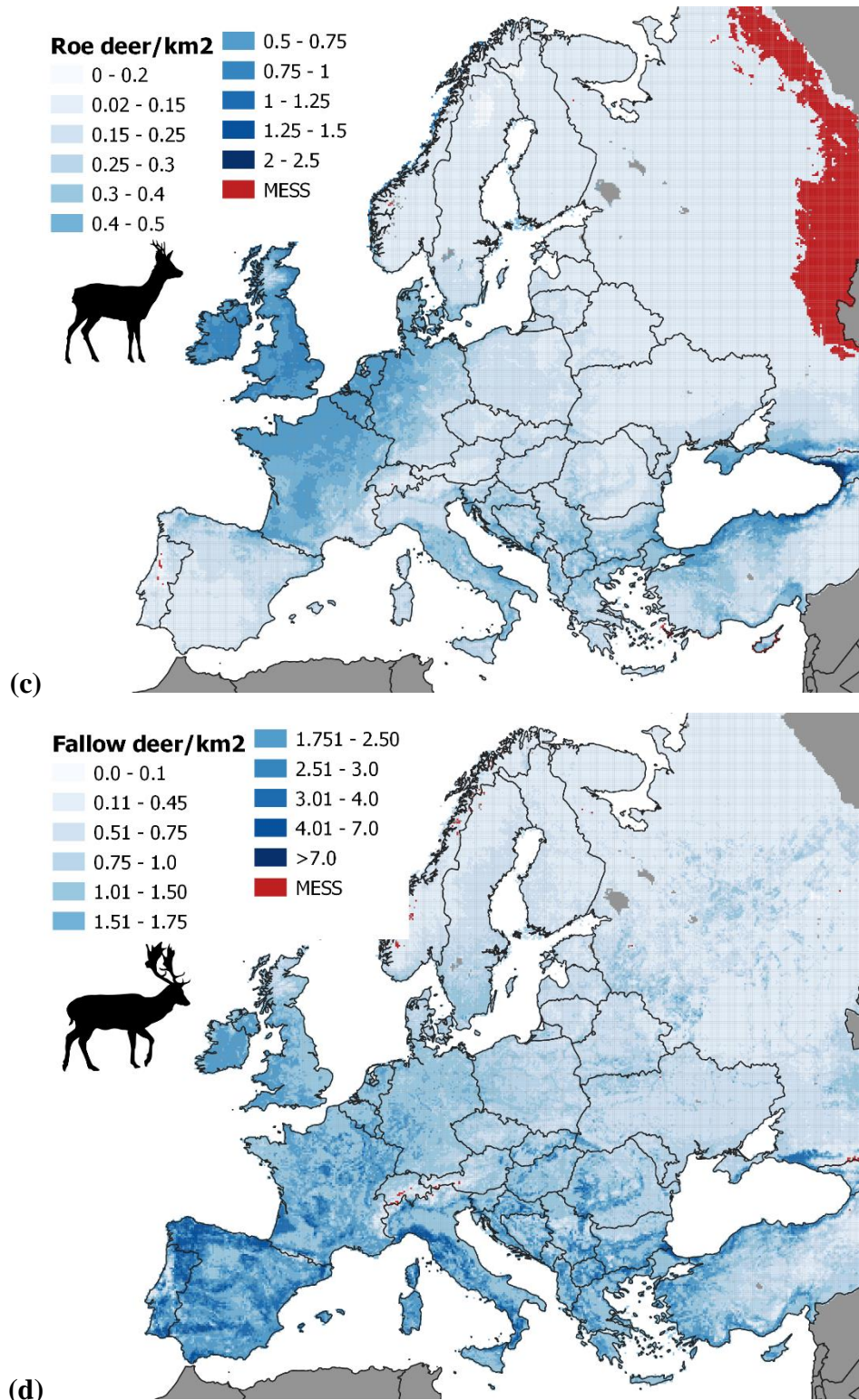
Overall, despite overprediction in several calibration plots, we found monotonicity in most cases. This indicates that models adequately represented the spatial pattern. Moreover, the removal of zeros on the validation datasets affected the calibration plots of all regions, showing a better predictive performance when zeros were removed. The patterns found (all data *vs* removing "0") could point out areas of potential expansion of the species according to environmental characteristics.

Contrary to widely distributed species, calibration plots for constrained species showed the importance of maintaining zero values in the datasets for a good model prediction. Also, if zero values are not considered, model predictions may have underestimated (Figure 9). The prediction of the models showed an accurate performance for all species but Northern chamois, reindeer,

and sika deer, in which predictive performance of the model neither show a clear pattern nor a good precise prediction. Due to the great overestimation of the predicted values for reindeer as well as its scarcity of data (Figure 9h), this species was not considered for modelling projections.

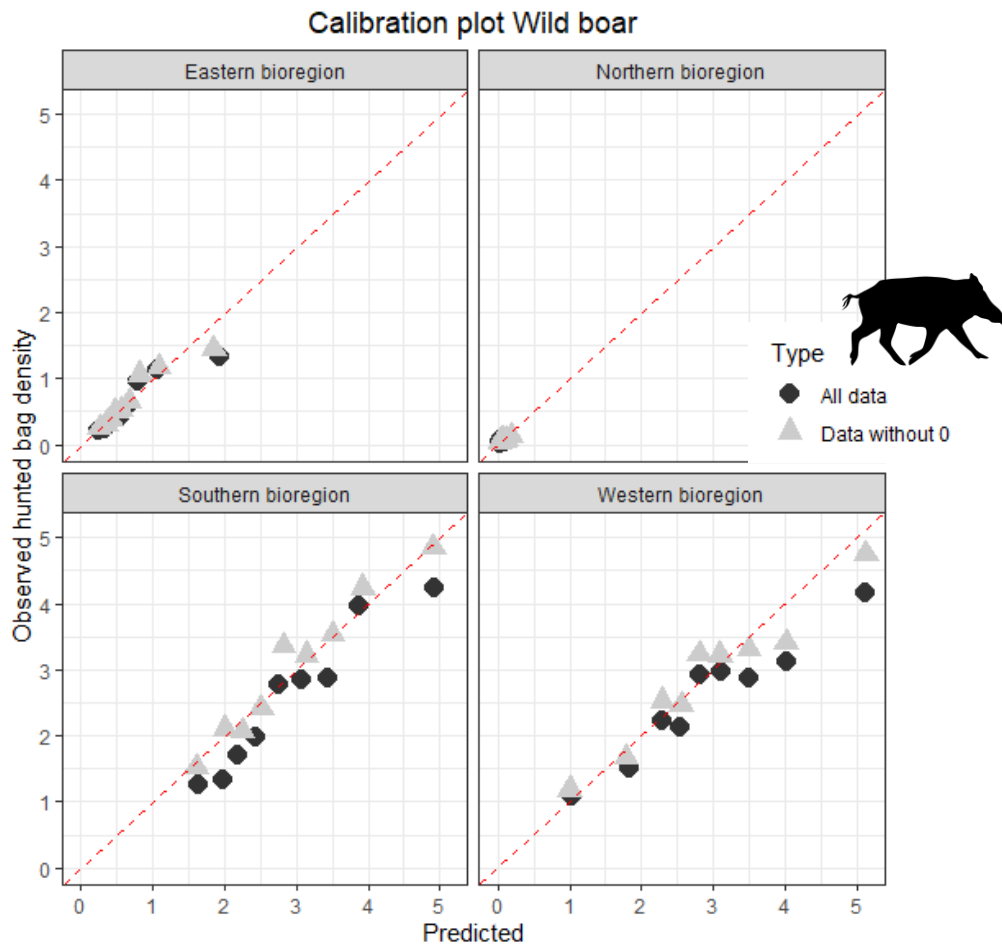




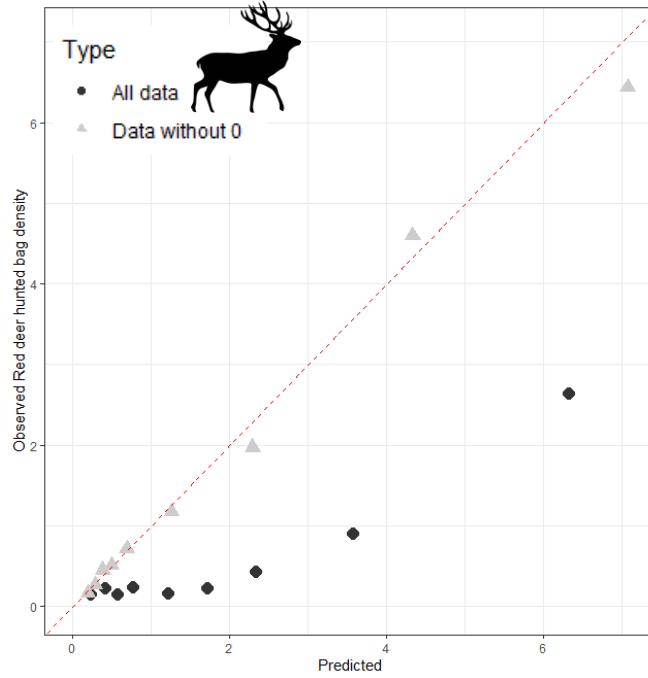


**Figure 6:** Hunting yield (HY) density (individual hunted per Km<sup>2</sup>) of widespread species: wild boar (a), red deer (b), roe deer (c) and fallow deer (d) at 10x10km. Red areas are beyond the environmental domain according to MESS analyses. The window on the right shows the bioregion classification for further discussion purposes.



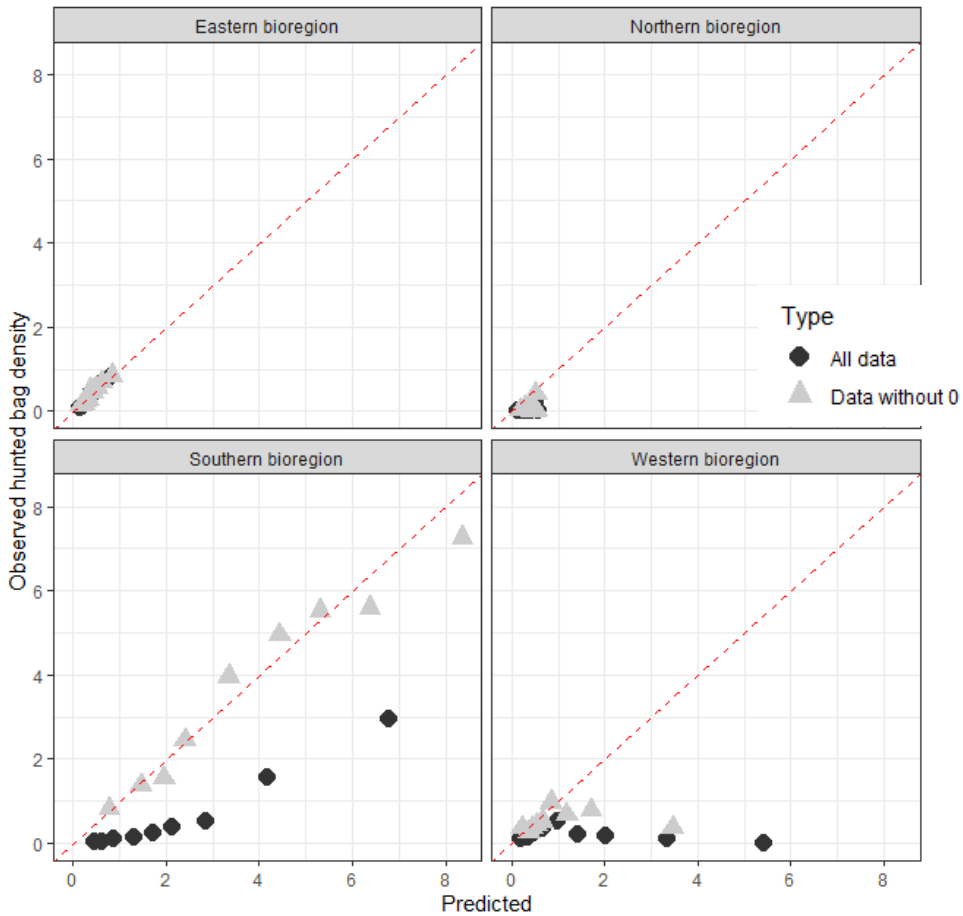


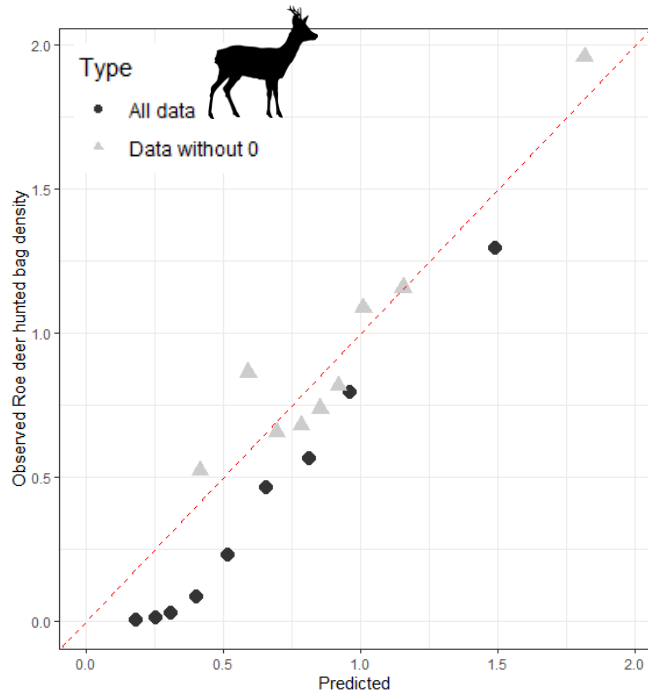
(a)



(b)

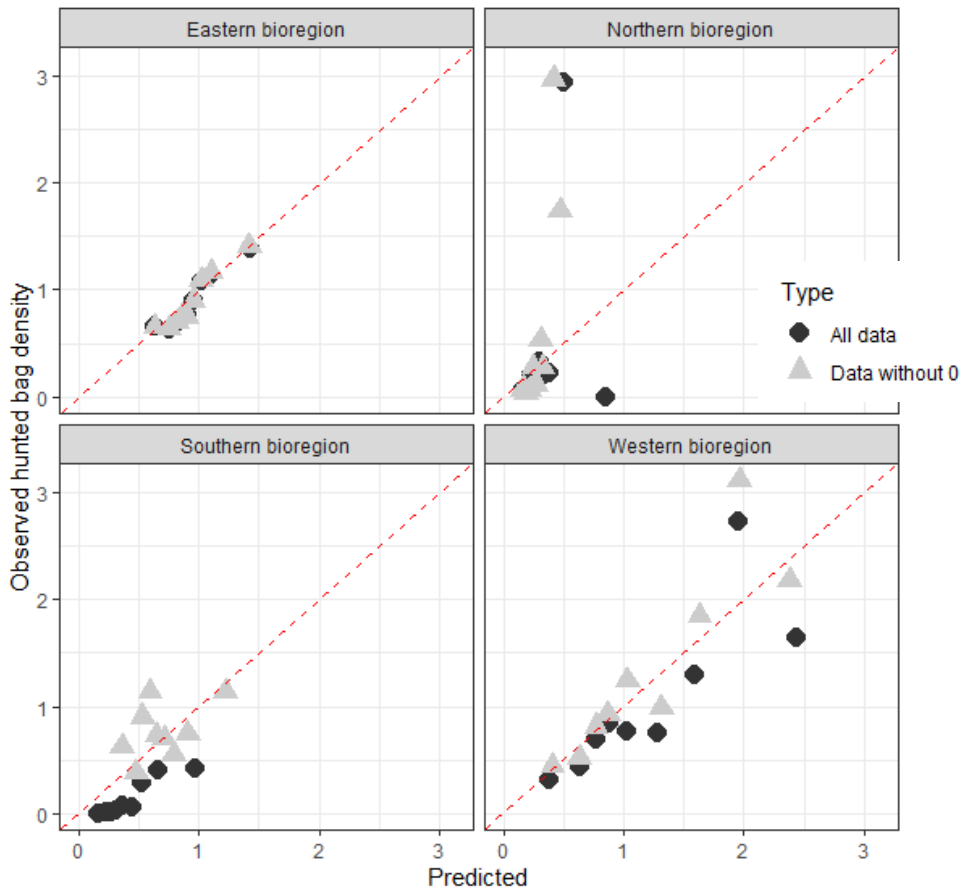
Calibration plot Red deer

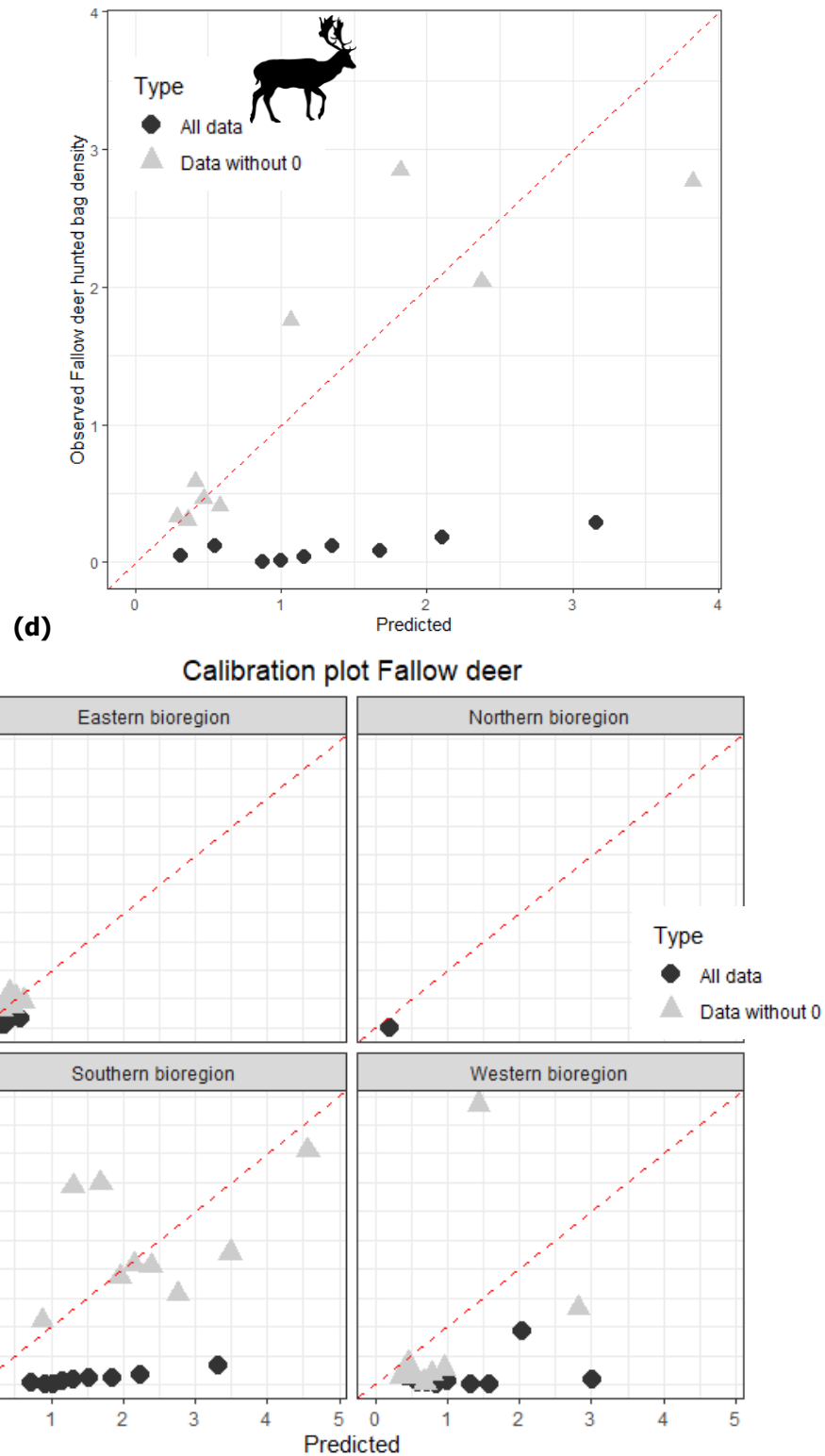




(c)

Calibration plot Roe deer





**Figure 7:** Calibration plot for assessing predictive performance of (a) wild boar HY model for each bioregion, (b) red deer HY model, for all Europe and by bioregions, (c) roe deer HY model, again for all Europe and by bioregions, and (d) fallow deer HY model, for all Europe and by bioregions. Plots show the relationship between the predicted hunting yield densities (HY) and the observed ones on the validation datasets.



HY models' predictions suggested different patterns depending on the species (Figures 6 and 8).

Spatial patterns of abundance (predicted HY) for wild boar are in agreement to previous reports (Figure 6a). Model projections showed an increase of HY density in Germany, probably as result of having better spatial data (lower resolution) than in previous reports in this country. A density gradient was found in the continent, reaching medium to high densities at Central-Western-South Europe. Densities drop off towards the North and East of Poland, showing a scattered pattern of medium to low densities in the Eastern and Northern regions. It is notable that moderate to high-density values predicted for southern Sweden and Norway, where climatic conditions are similar to Western bioregion and model projection must be influenced by it.

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 in *ENETWILD* consortium et al. (2022) decreased in this model (Figure 6b) as it was expected due to the more restricted distribution of the species in this bioregion.

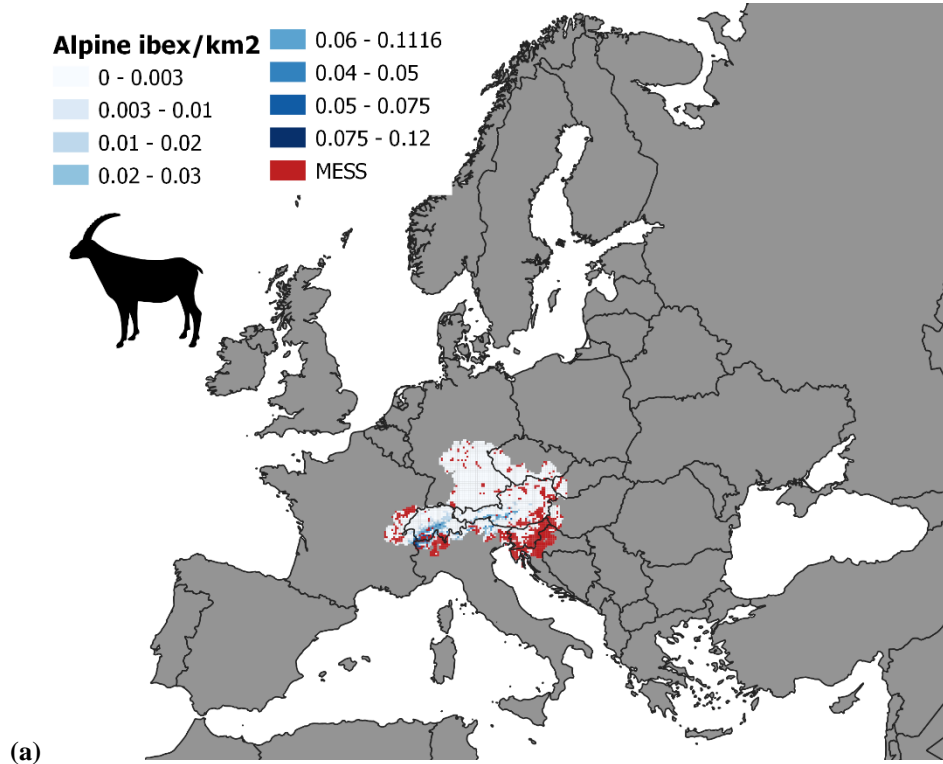
Concerning the roe deer model, the highest HY is found in Atlantic temperate area (including British Isles), progressively decreasing towards North (Northern region), East (Eastern region) and South (Mediterranean region), respectively. Overall, the calibration plot performed better than for red deer. Similarly, a low correspondence was found in the Northern region (Figure 6c), which may also be due to the low data availability in this bioregion.

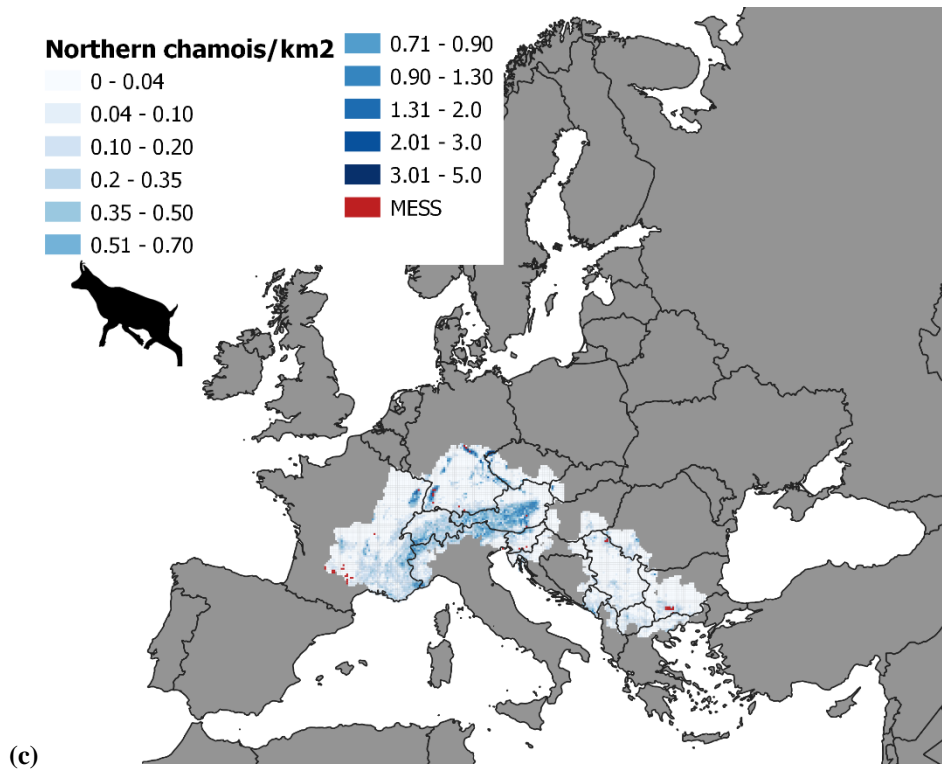
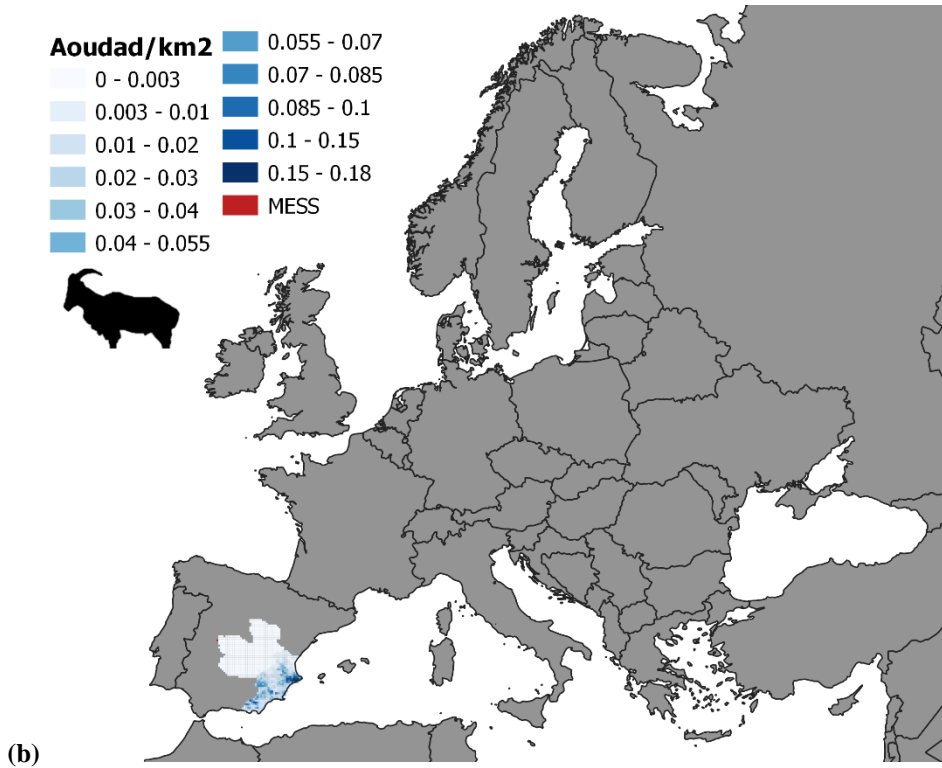
Special caution must be taken with spatial predictions of fallow deer as they mostly showed higher potential HY density than real expected. Although it can be considered a widespread species in Europe, their populations are normally locally distributed and may have not yet achieved the ecological equilibrium in the continent. According to the results, highest HY densities of fallow deer will be expected in the Iberian Peninsula, which will progressively decrease towards North and East of Europe (Figure 6d).

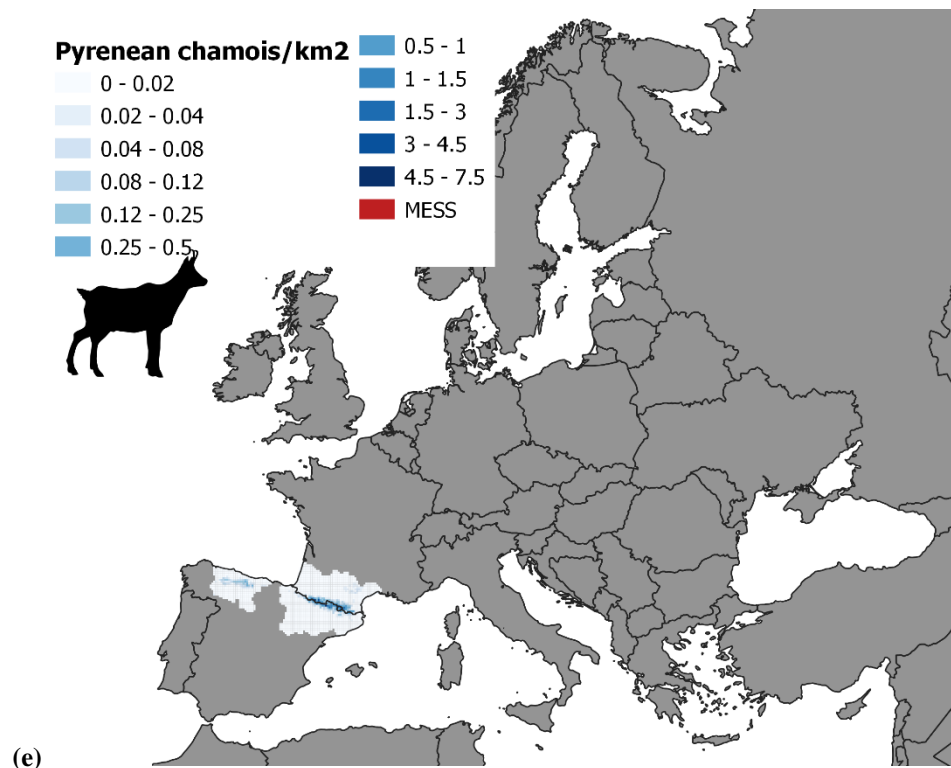
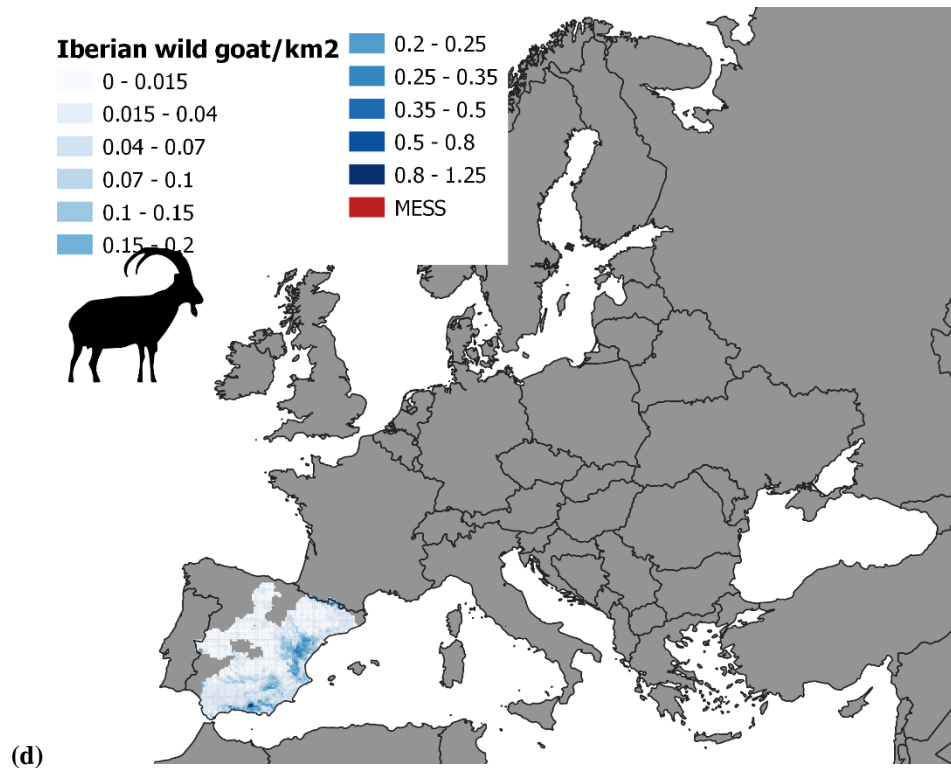
Overall, our results are consistent with the expected abundance distribution of wild boar, red deer, European roe deer and fallow deer. They are the most common ungulate species in Europe and can occupy a diversity of habitats, including deciduous and coniferous continental forests, Mediterranean scrublands, agricultural plains (the latter mainly applies to roe deer), but also high latitudes and altitudes, where harsh winter occur. Each species showed a different spatial pattern as well as different values of HY densities, with red deer and wild boar the species with higher HY densities in Europe, followed by the potential HY density of fallow deer, and, finally, by roe deer, which is less ubiquitous in Mediterranean areas, and finds its optimal in Atlantic habitats, which is well captured by our model.

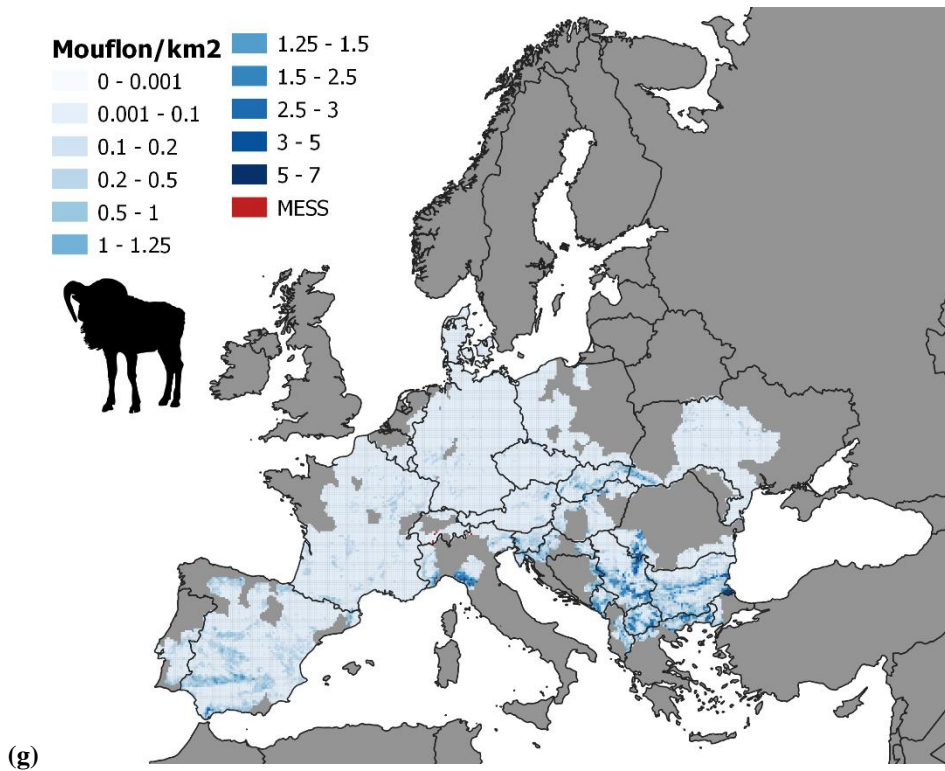
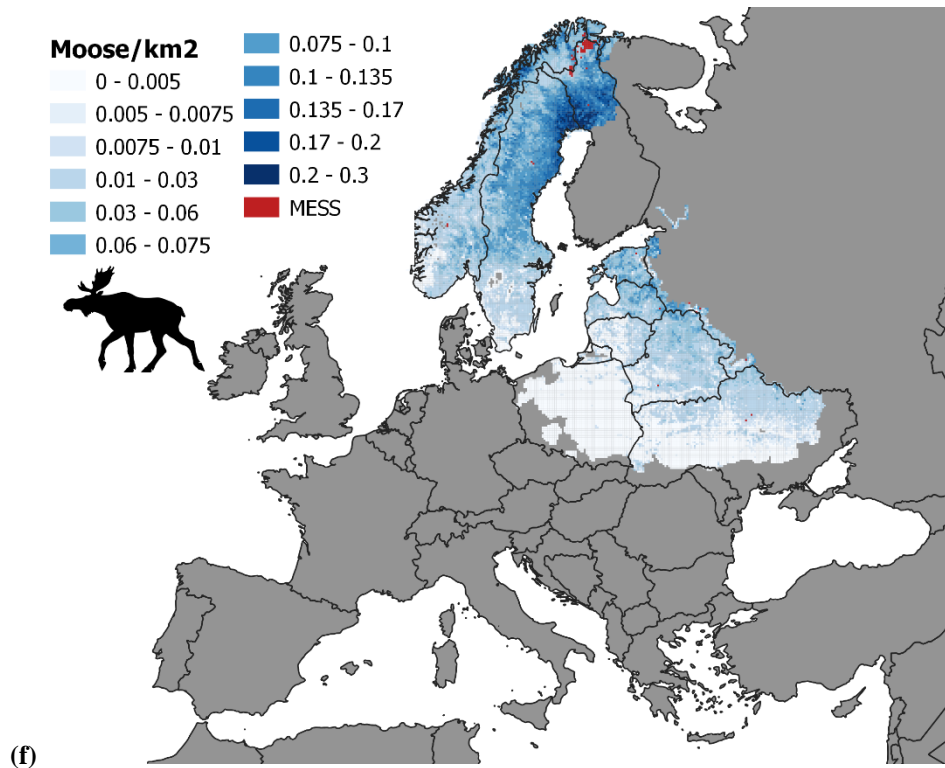
Spatial patterns for constrained distributed species showed a reliable abundance distribution for most species. Alpine ibex showed higher abundances where expected, around the Alps mountains (Figure 8a). The HY abundance distribution pattern agreed with the ecological preference for aoudad (Figure 8b) and for chamois. For the latter, higher abundances are found in the Alps according to its known adaptation to the Alpine environment (Brivio et al, 2016), showing a lower abundance pattern around Serbia, but with an increase of abundance values in North Macedonia and Bulgaria, which is probably due to the influence of high mountain chains in those countries (Figure 8c). The spatial abundance pattern of Iberian wild goat was also reliable, although its abundance in Central Spain was missing according to this projection, probable due to the lack of data reported for this species in those places (Figure 8d). The distribution pattern of the Pyrenean chamois agreed to the expected higher values associated to the Iberian Mountain chains (Figure 8e). The moose abundance HY pattern showed a West-North-eastern gradient with show higher values at higher latitudes. It showed a good spatial pattern in the Baltic countries, but Norway and Sweden were expected to show abundance values higher at lower latitudes in the Central-South region of both countries (Figure 8f). On the contrary, neither mouflon nor sika deer showed

a reliable pattern of abundance distribution. In the first case, there was a reliable pattern for the Iberian Peninsula, but it probably did not show the abundance distribution pattern for Germany, Czech Republic, and Slovakia where higher values were expected, nor for Serbia, and North Macedonia, where lower values were expected (Figure 8g). In the second case, we expected higher values of abundance in Central Europe (Germany, Austria, and Poland) in comparison to those obtained for Ukraine and France (Figure 8h).

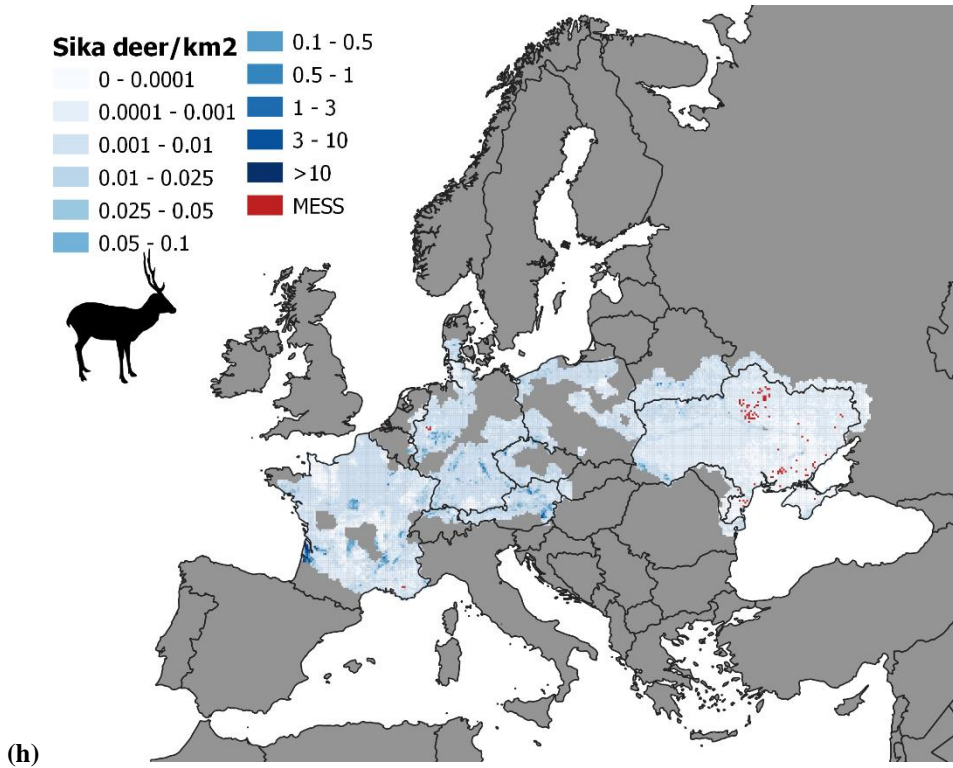




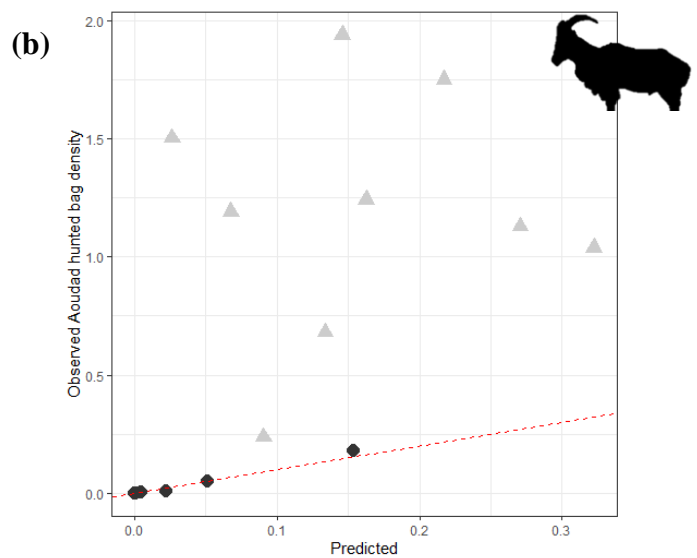
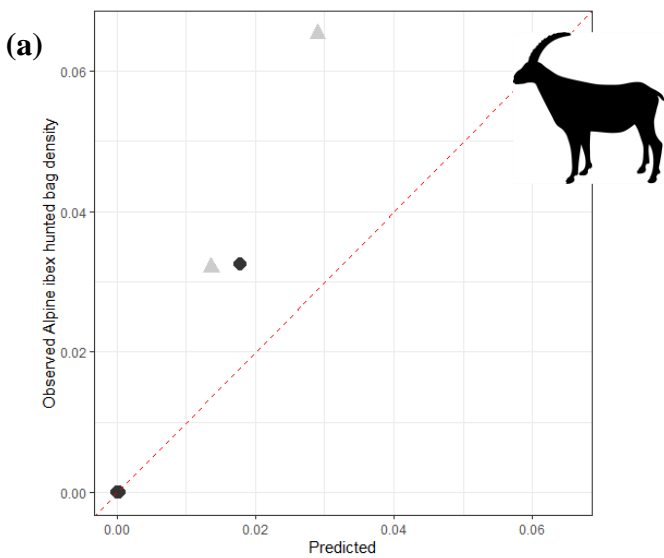


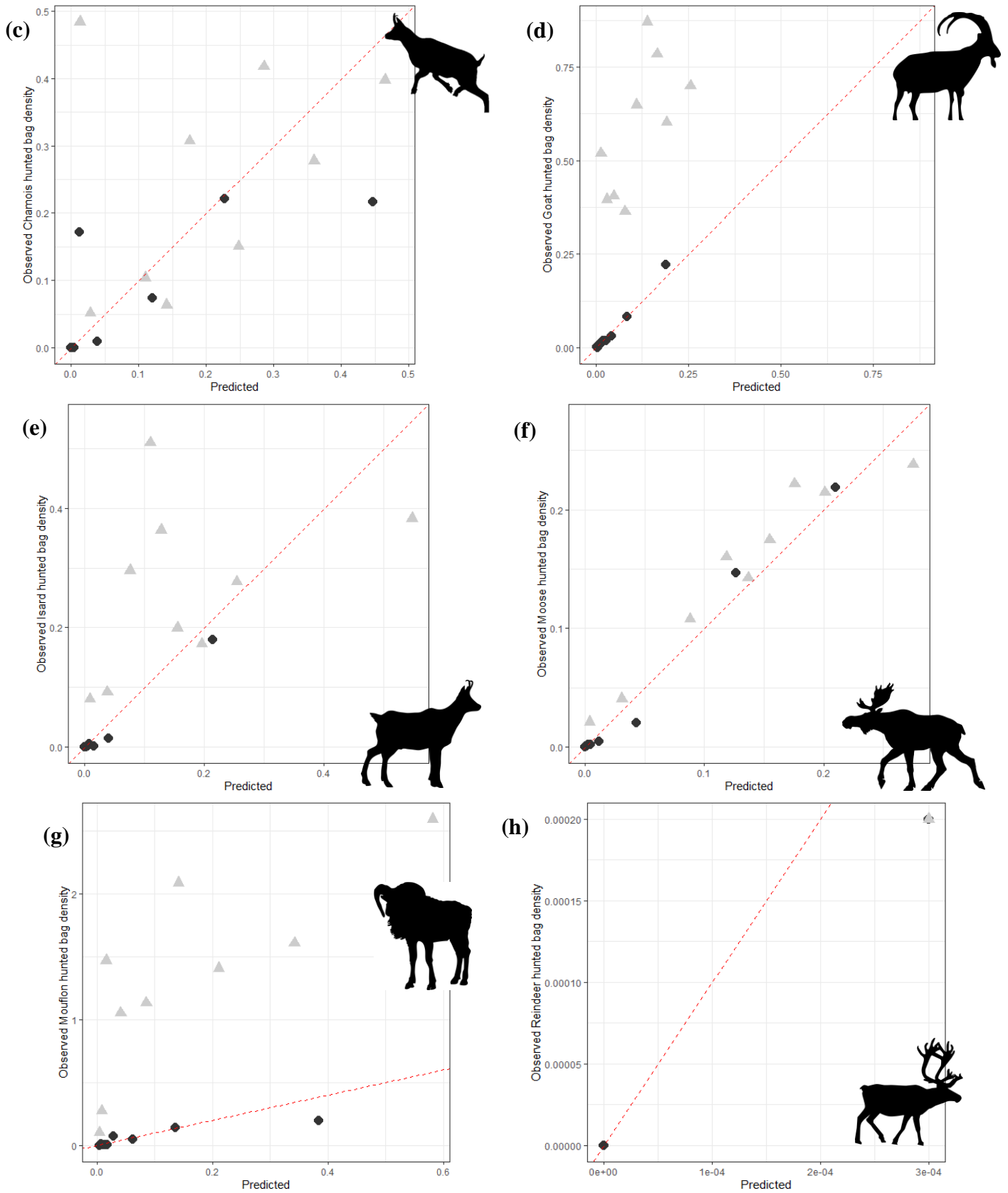


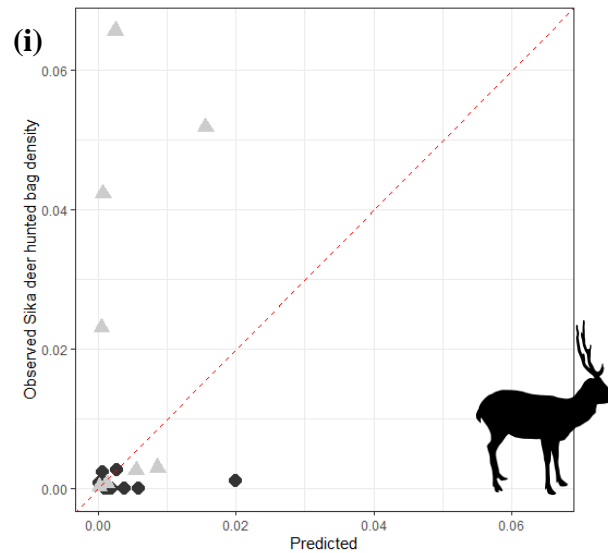




**Figure 8:** Hunting yield (HY) density (individual hunted per Km<sup>2</sup>) of constrained species: Alpine ibex (a), aoudad (b), Northern chamois (c), Iberian wild goat (d), Pyrenean chamois (e), moose (f) mouflon (g), and sika deer (h) at 10x10km. Red areas are beyond the environmental domain according to MESS analyses.





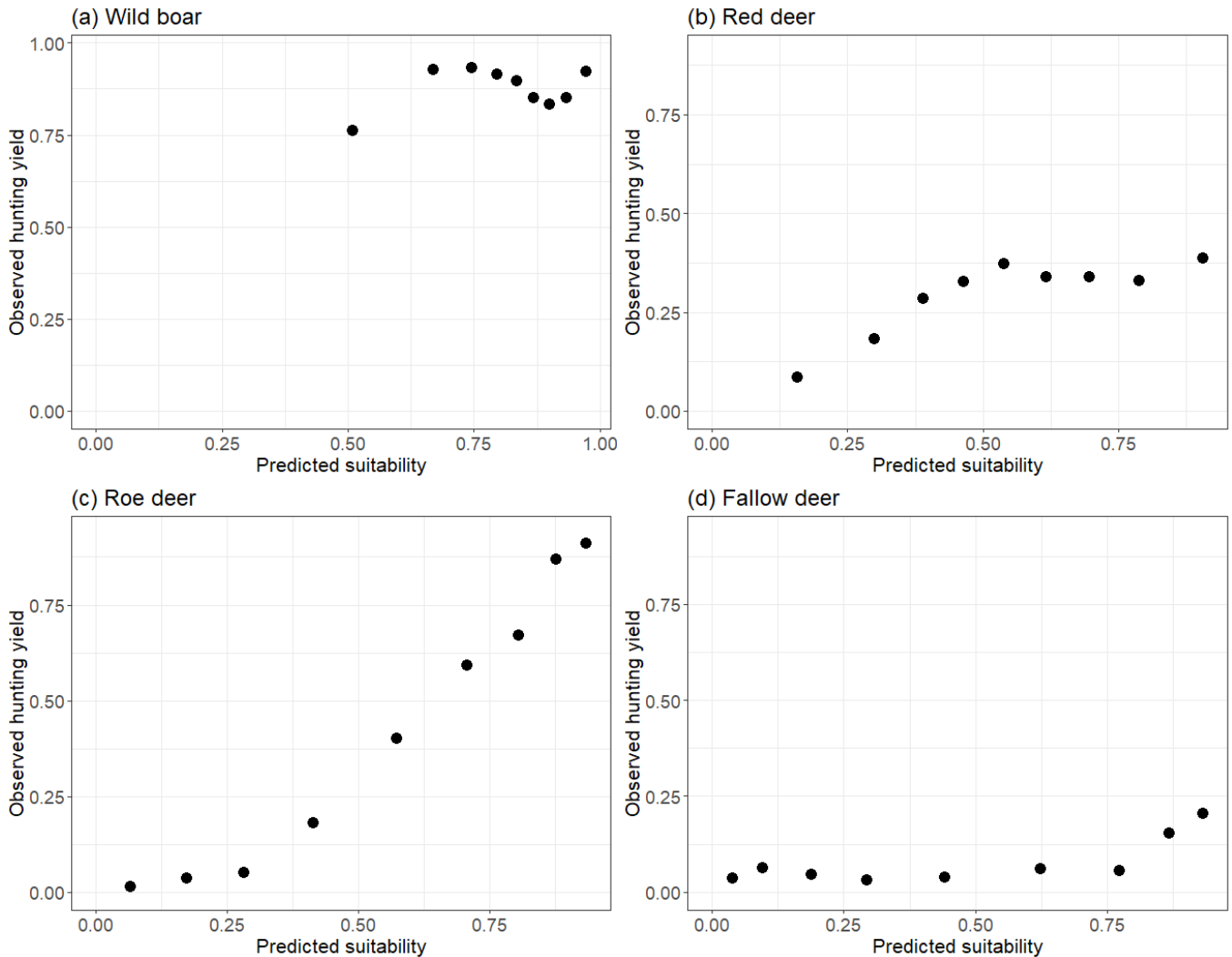


**Figure 9:** Calibration plot for assessing predictive performance of (a) Alpine ibex, (b) aoudad, (c) Northern chamois, (d) Iberian wild goat, (e) Pyrenean chamois, (f) moose, (g) mouflon, (h) reindeer, and (i) sika deer HY model, for all Europe. Plots show the relationship between the predicted hunting yield densities (HY) and the observed ones on the validation datasets.

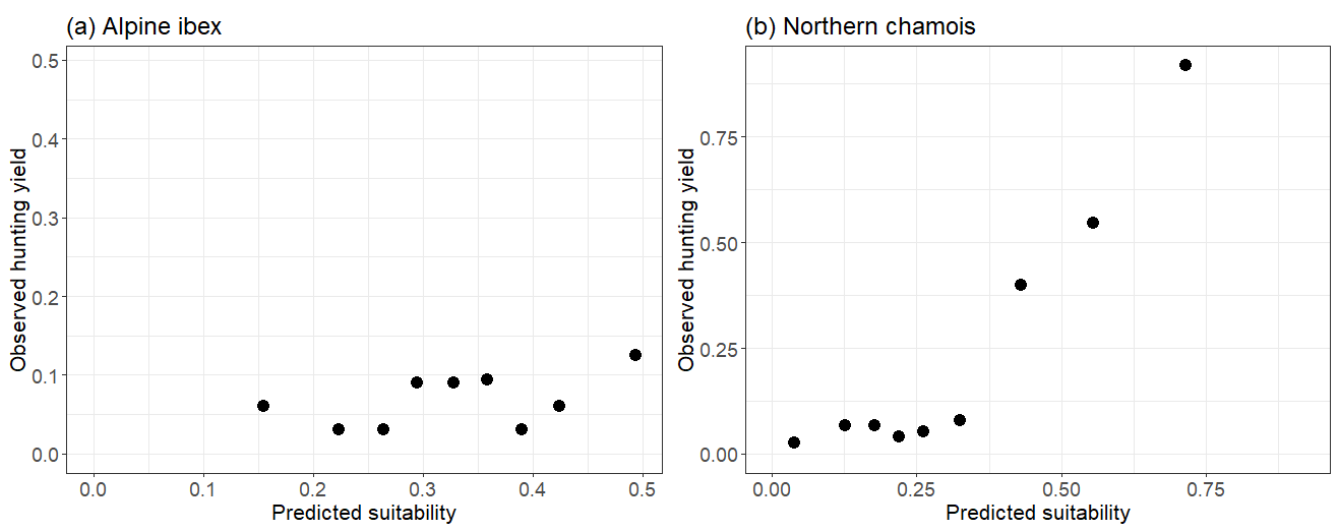
### 4.3. Comparison of models based on hunting yield and occurrence data

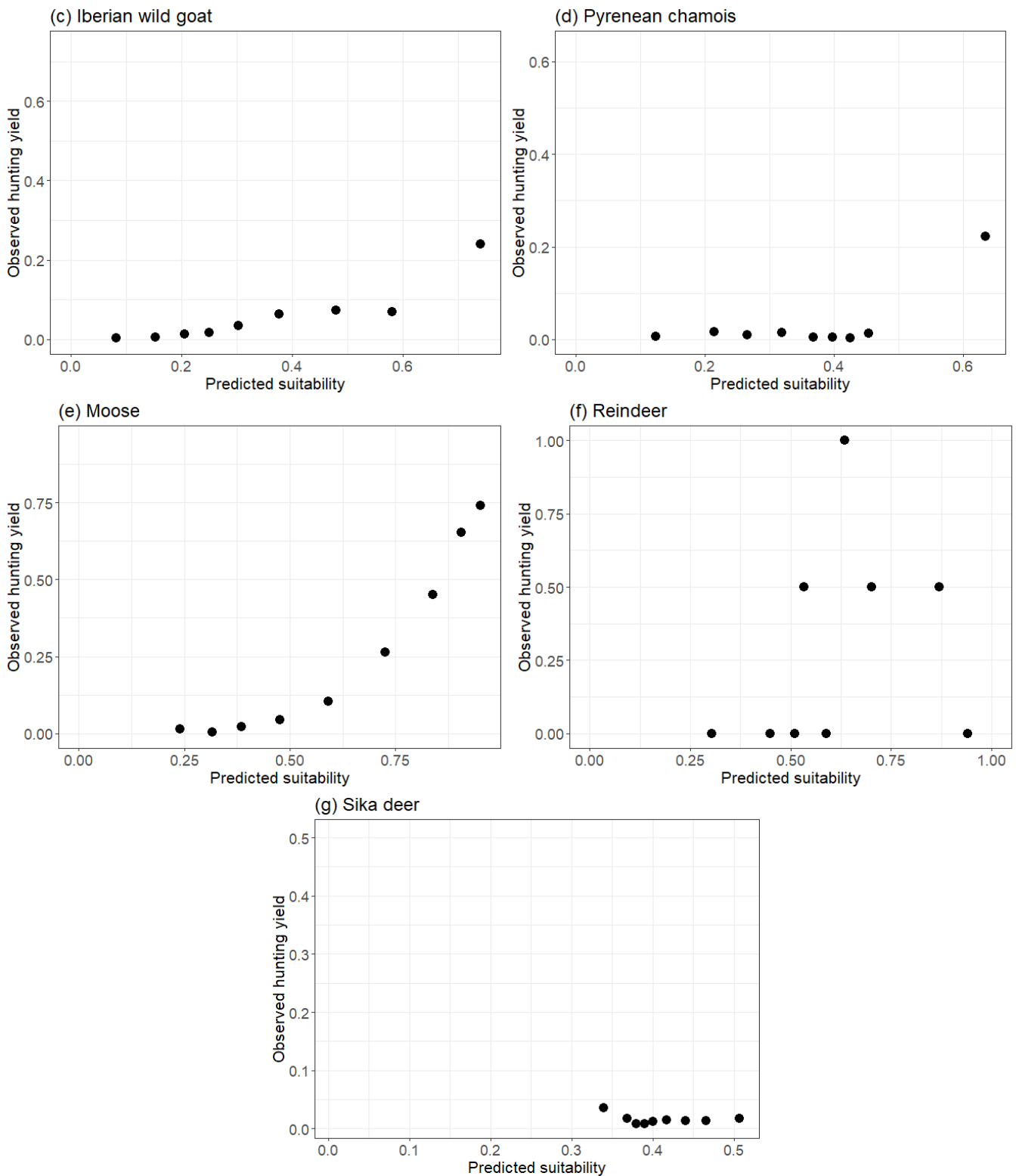
The observed HY and habitat suitability showed a flat pattern for wild boar, indicating that the wild boar hunting yield remained constant across habitat suitability values (Figure 10a). Moreover, red deer showed a linear pattern where observed HY increased with higher suitability values until medium values of habitat suitability, where the linear slope became flat. In other words, suitability can be considered as a proxy of HY, although as a proxy, it would have low resolution as it is only able to distinguish between four HY classes (Figure 10b). Furthermore, roe deer showed a good pattern, being able to differentiate more classes, especially in the last intervals, which means that spatial pattern for localities with high suitability (above 0.8) could properly distinguish observed HY (Figure 10c). Finally, fallow deer also showed a linear pattern, although the relation is almost flat for first classes indicating that first intervals classes would belong to a same proxy index, while at higher suitability predicted values (above 0.75) two observed HY classes of fallow deer could be differentiated (Figure 10d).

For constrained distributed species, the patterns showed linear relationship for more than the half of the comparisons made, indicating that predicted suitability could be used as a proxy of observed HY. However, the resolution of the intervals would depend on each species. Species such as moose would have a high-resolution interval, being able to differentiate between all the bins (Figure 11e); Northern chamois could distinguish less classes and last interval classes could be more precise for determining observed HY (Figure 11b). Species as Iberian wild goat, and Pyrenean chamois could differentiate two intervals: one with low HY values and another in which HY values sharply incremented at suitability values above 0.6 (Figures 11c, d), while sika deer showed the contrary pattern, that is, higher values of HY at lower values of suitability and a decline on observed HY when suitability values increased (Figure 11g). Finally, no relationship was found neither between Alpine ibex nor reindeer hunting bags and habitat suitability values (Figures 11a and f).



**Figure 10:** Calibration plots of widespread species: (a) wild boar, (b) red deer, (c) roe deer, and (d) fallow deer showing the relationship between predicted suitability and observed hunting yields values.





**Figure 11:** Calibration plots of constrained species: (a) Alpine ibex, (b) Northern chamois, (c) Iberian wild goat, (d) Pyrenean chamois, (e) moose, (f) reindeer, and (g) sika deer showing the relationship between predicted suitability and observed hunting yields values.



#### 4.4. Conclusions and further steps

##### *Occurrence data model*

- A two-step occurrence model, identifying likely absence, and then exploring powerful machine learning algorithms to fit presence-absence data, yields improved predictions across all species compared with the previous coupled Bayesian framework. There are several plausible explanations for this, but one factor may be the ability of random forest to better capture non-linear relationships between occurrence (suitability) and environmental variables.
- The new approach allows the use of raw variables (rather than requiring an opaque PCA transformation to facilitate fitting; ENETWILD consortium et al., 2021). While random forest is somewhat of a “black-box” inference on variable importance, which could provide valuable insight/verification. Such outputs can however be influenced by factors like co-correlation (Strob et al., 2008). Further work is required to understand these issues before robust inference can be gained but potential solutions are available (party package in R; Strob et al., 2008).
- Identification of absence prior to modelling allows examination of the spatial structure and balance between presences and absences. It has been suggested that these patterns can impact model outputs, for instance where mixing is poor. Additional investigation is required to determine if manipulation (careful subsampling) of presence and absence locations could yield further improvement in model outputs (Steen et al., 2021).

##### *Hunting yield density data model*

- Model projections showed good abundance patterns for most species. There is not a single framework for modelling at the European scale and it should be adapted to the particularities of the distribution of the dataset for modelling. The three frameworks proposed were a good approximation for modelling the abundance distribution of species.
- The zeros, i.e., the areas where the species is not hunted, are not informative for modelling species with wide distribution range and therefore should be removed from the dataset but they should be maintained when modelling species with constrained ranges.
- A patchy distribution of collected HY lead to a weird spatial pattern of abundance in several species (e.g., sika deer).
- For some species, there is still a need for better quality of data (amount and better spatial resolution if available) for modelling purposes (e.g., reindeer).

##### *Validation of suitability on HY*

- The comparison of observed HY and the prediction of the models based on occurrence indicated that suitability predicted values agree with hunting yields values for most species. This may indicate that suitability has potential to be used as a proxy for abundance in this species. Moreover, the resolution of intervals for using suitability as a proxy of abundance will depend on each species. However, habitat suitability did not show potential to be used as a proxy for few species.
- The different relationships seen between species may suggest that incorporating habitat suitability values in the HY models as explanatory variable may not always increase the explained variance of the model and that it is necessary to clarify the relationship and determine if its inclusion could benefit HY modelling results in all species.

## 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.
- Brivio F, Bertolucci C, Tettamanti F, Filli F, Apollonio M, Grignolio S, 2016. The weather dictates the rhythms: Alpine chamois activity is well adapted to ecological conditions. *Behavioral Ecology and Sociobiology*, 70, 1291-1304.
- 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-Aguilar 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-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, 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.
- ENETWILD-consortium, Illanas S, Croft S, Acevedo P, Fernández-López J, Vicente J, Blanco-Aguilar J A, 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. . EFSA Supporting Publication 2022:EN-7174.
- Hattab T, Garzón-López CX, Ewald M, et al., 2017. A unified framework to model the potential and realized distributions of invasive species within the invaded range. *Diversity and Distributions*, 23: 806-819.
- 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.
- 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.
- Legendre P, Fortin MJ, 1989. Spatial pattern and ecological analysis. *Vegetatio*, 80: 107-138.
- Liaw A, Wiener M, 2002. Classification and Regression by random Forest. *R News*, 2(3): 18-22.
- 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.
- Maes D, Isaac NJ, Harrower CA, Collen B, Strien AJ, Roy DB, 2015. The use of opportunistic data for IUCN Red List assessments. *Biological Journal of The Linnean Society*, 115: 690-706.

- 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.
- Naimi B, Hamm Na, Groen TA, Skidmore AK, Toxopeus AG, 2014. Where is positional uncertainty a problem for species distribution modelling. *Ecography*, 37: 191-203.
- 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.
- Ridout M, Hinde J, Demétrio C G B, 2001. A score test for testing Zero-Inflated Poisson regression model against Zero-Inflated Negative Binomial alternatives. *Biometrics*, 57:219-223.
- Royle JA, 2006. Site Occupancy Models with Heterogeneous Detection Probabilities. *Biometrics*, 62: 97-102.
- Steen VA, Tingley MW, Paton PWC, Elphick CS, 2021. Spatial thinning and class balancing: Key choices lead to variation in the performance of species distribution models with citizen science data. *Methods in Ecology and Evolution*, 12: 216-226
- Strobl C, Boulesteix AL, Kneib T, Augustin T, Zeileis A, 2008. Conditional variable importance for random forests. *BMC Bioinformatics*, 9: 307.
- 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

<b>Absolute population density</b>	Number of individuals per surface unit, usually by km <sup>2</sup> . This is an absolute measure that allows to make direct comparison among populations.
<b>Abundance estimate</b>	The number of individuals in a population calculated by statistical methods.
<b>ASF</b>	African Swine Fever.
<b>AUC</b>	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.
<b>Bayesian hierarchical framework</b>	Statistical model written in multiple levels (hierarchical form) that estimates the parameters of the posterior distribution using the Bayesian method
<b>Bioregion</b>	Homogeneous bioclimatic regions based on bioclimatic variables, vegetation cover and topographic covariates associated to wild boar density
<b>Cross-validation</b>	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.
<b>Downscaling</b>	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.
<b>Environmental domain</b>	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.
<b>GBIF</b>	Global Biodiversity Information Facility.
<b>GLM</b>	Generalised Linear Model.
<b>Habitat suitability</b>	The ability of a habitat to provide a species life requisites under current conditions.
<b>Hunting bag</b>	It refers to the number of animals hunted in a territory usually during a given hunting season.
<b>Hunting yield</b>	It is usually used to refer to a relative abundance index based on hunting bag data.
<b>IUCN</b>	International Union for Conservation of Nature.
<b>MESS</b>	Multivariate Environmental Similarity Surface.
<b>Model extent</b>	This term refers to the geographical area on which the model is to be fitted.
<b>NUTS3</b>	Nomenclature of Territorial Units for Statistics Level 3.



<b>Population density (d)</b>	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).
<b>Population size or absolute abundance (N)</b>	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.
<b>Predictive accuracy</b>	Quantitative metric describing the accuracy of model predictions. Computed by comparing model predictions against independent data often obtained through a process of cross-validation.
<b>Presence-absence</b>	Datasets contain independent locations with binary classification describing whether a species is present or explicitly absent.
<b>Presence-background</b>	Independent datasets describing environmental conditions at locations where a species has been observed and those of a random sample from the available landscape.
<b>Presence-only</b>	Dataset containing independent events (date, location, recorder) describing species sightings; positive occurrences.
<b>Random forest</b>	Modelling algorithm based on regression trees.
<b>Relative abundance</b>	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.
<b>Relative score</b>	Index describing the difference in suitability, i.e., likelihood of species presence, across locations.
<b>RSF</b>	Resource Selection Function.
<b>Suitability</b>	Measure of how suitable a location is for a particular species; analogous to the likelihood that a species is present.
<b>Training dataset</b>	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).
<b>TSS</b>	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.
<b>Variable importance</b>	Quantitative measure of the relative importance/contribution of model variables in explaining observed data.