



EXTERNAL SCIENTIFIC REPORT

APPROVED: 20 February 2023

doi:10.2903/sp.efsa.2023.EN-7894

Wild carnivore occurrence and models of hunting yield abundance at European scale: first models for red fox and badger

ENETWILD-consortium¹, Sonia Illanas, Simon Croft, Graham C. Smith, Joaquín Vicente, José Antonio Blanco-Aguiar, Massimo Scandura, Marco Apollonio, Ezio Ferroglio, Oliver Keuling, Kamila Plis, Sándor Csányi, Azahara Gómez-Molina, Ludovica Preite, Carmen Ruiz-Rodríguez, Sergio López-Padilla, Stefania Zanet, Rachele Vada, Tomasz Podgorski, Francesca Brivio, Javier Fernández-López, Ramón C. Soriguer, Pelayo Acevedo

Abstract

The goal of this report is modelling the occurrence for carnivores at the European scale and to compare the output of occurrence with observed hunting yield (HY) density models for red fox (*Vulpes vulpes*) and badger (*Meles meles*). Random Forest function was used for modelling occurrence of species. Occurrences available from the past 30 years (1990-2020), and HY data (period 2012-2021) from records submitted to ENETWILD were considered for modelling. Like previous models based on HY for ungulates, the response variable was the maximum number of carnivores hunted in that period divided by the area in km² of the corresponding administrative unit (HY density). Models based on HY were statistically downscaled to make predictions to 10x10 km². Occurrence data models indicated a good predictive performance for most species, showing that the model framework proposed for ungulates can also be applied for carnivores. Realistic distribution maps of carnivore species were achieved under this framework, except for those ones which are expanding their range, the golden jackal (*Canis aureus*), or those considered alien species, raccoon (*Procyon lotor*) and raccoon dog (*Nyctereutes procyonoides*); or those having a very limited distribution as the Iberian lynx (*Lynx pardinus*) or the steppe polecat (*Mustela eversmannii*): in those cases the obtained models were underestimating their suitability in Europe. Suitability has potential to be used as a proxy for abundance of red fox and badger. Validation of suitability on HY suggested the potential to be used as a proxy for abundance of red fox and badger but depending on each species. The calibration plots for HY models showed a good and linear predictive performance for fox and badger as well as an expected pattern of abundance of species, according to the data. However, differences in type of hunting and regulations in game carnivores between countries must be playing an important role in the patterns obtained. We conclude that (i) the framework developed for modelling ungulates distribution generally well fit to carnivores species, (ii) the predicted suitability were realistic for all carnivores, but alien invasive species, limited distributed species and species expanding its range, and (iii) HY model projections displayed good abundance patterns for red fox and badger, showing that the frameworks proposed for wild ungulates were a good approximation for modelling the distribution of carnivores HY. As a future step, we need to explore how to improve the results when the unavailability of hunting activity for some species limits the extrapolation to other regions.

© European Food Safety Authority, 2023

Key words: distribution, hunting bags, occurrence, population abundance, population monitoring, risk assessment, spatial modelling, carnivores, red fox, badger

Question number: EFSA-Q-2022-00046

Correspondence: biohaw@efsa.europa.eu

¹ ENETWILD Consortium: 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 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 wildlife.

Suggested citation: ENETWILD-consortium, Sonia Illanas, Simon Croft, Graham C. Smith, Joaquín Vicente, José Antonio Blanco-Aguiar, Massimo Scandura, Marco Apollonio, Ezio Ferroglio, Oliver Keuling, Kamila Plis, Sándor Csányi, Azahara Gómez-Molina, Ludovica Preite, Carmen Ruiz-Rodríguez, Sergio López-Padilla, Stefania Zanet, Rachele Vada, Tomasz Podgorski, Francesca Brivio, Javier Fernández-López, Ramón C. Soriguer, Pelayo Acevedo, 2023. Wild carnivore occurrence and models of hunting yield abundance at European scale: first models for red fox and badger. EFSA supporting publication 2023:EN-7894. 26 pp. doi: 10.2903/sp.efsa.2023.EN-7894

ISSN: 2397-8325

© European Food Safety Authority, 2023

Reproduction is authorised provided the source is acknowledged.

Summary

Background and objectives: Recently the *ENETWILD* consortium has been developing generic model frameworks for predicting habitat suitability and hunting yield densities (HY) first for wild boar, as a study case, and afterwards for widely and constrained distributed ruminant species across Europe. Overall, the model frameworks developed well for these ungulates' species.

The goal of this report is to apply the models developed for ungulates to carnivore's species and to explore if these approaches can be used (i) for modelling the occurrence at European scale, (ii) implementing HY density model framework for red fox and badger, and (iii) comparing the output of occurrence with observed HY.

Data: We used occurrences available from the past 30 years (1990- 2020) through the Global Biodiversity Information Facility (GBIF) extracted on 04/11/2022² 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 2012-2021 from records submitted to *ENETWILD* Data Model, extracted on 05/12/2022.

Modelling: Random Forest function was used for modelling occurrence of species. Regarding HY based models, the response variable was the maximum number of individuals annually hunted in 2012-2021 hunting seasons divided by the area (km²) of the corresponding administrative unit (HY density). We conducted GLM negative binomial models for red fox and GLMM (mixed models) with negative binomial distribution for badger, using country as random effect. All models, occurrence and HYs, included eco-geographical variables as predictors (fixed effects). 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 indicated a good predictive performance for most species exceeding 0.7 AUC value for thirteen of the sixteen species modelled, showing that the model framework proposed for ungulates can also be properly applied for carnivores. Realistic distribution maps of carnivore species were achieved under this framework, except for those which are expanding their range such as the golden jackal (*Canis aureus*), are considered alien species such as raccoon (*Procyon lotor*) and raccoon dog (*Nyctereutes procyonoides*) or have a very limited distribution such as the Iberian lynx (*Lynx pardinus*) or the steppe polecat (*Mustela eversmannii*), in these cases the distribution maps obtained were underestimating their suitable range in Europe. The validation with hunting yield showed that suitability has potential to be used, specifically generated for each specie, as a proxy for abundance of red fox and badgers. On the other hand, calibration plots for HY models showed a good and linear predictive performance for both species as well as an expected pattern of abundance of species, according to the data. However, differences in hunting modalities and regulations in game carnivores between countries must be playing an important role in the patterns obtained, being limited to use hunting bags as comparison abundance index for the different countries, while hunting bags could be used as abundance index for comparing patterns within the same country for time series.

Conclusions and next steps:

Occurrence data model

- We have previously highlighted the potential use of variable importance as an additional tool for model validation alongside statistical metrics of performance. While such information is

² <https://doi.org/10.15468/dl.c7v9a6>

simple to extract from models it's interpretation can be more difficult and further work is ongoing to ensure any inference is robust.

- As part of the modelling approach, for simplification, we assume that observability is constant across our extent of interest. However, several factors may confound this assumption: substantial variation in species richness / behaviour (ecological) and recorder behaviour (anthropological). Additional work is required to assess the extent of any variation and the impact of any bias it may introduce in model predictions.

Hunting yield density data model

- The difference in hunting traditions and conservation status within each country on carnivore species must be considered for modelling carnivores' HY.
- The framework proposed for widely distributed ungulates seemed to be a good approximation and consistent for modelling widely distributed and hunted carnivores' species as it is the red fox, although results showed a slight overprediction.
- A new framework proposed for badger seemed to be a good approximation for modelling the abundance distribution of species that may have different legislation and hunting tradition between countries.
- Regarding the good results obtained with the new framework developed for badger, red fox models could consider implementing this framework, and to evaluate if results improve in comparison with the widely distributed ungulates framework.
- Model projections showed, in general, good abundance patterns for red fox and badger. There was not a single framework for modelling carnivores at the European scale and it should be adapted to the particularities of the distribution of the dataset for modelling. Moreover, using or removing zeros from the datasets for HY modelling was important and should be considered when modelling.
- It should be explored how to improve the results and projection of the HY models when the lack of hunting activity for some species in some locations limits the extrapolation to other regions.

Validation of suitability on HY

- Suitability has potential to be used as a proxy for abundance of red fox and badger. As it happened for ungulate species, the resolution of intervals for using suitability as a proxy of abundance will depend on each 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. Environmental variables and other predictors | 6 |
| 2. Data | 8 |
| 2.1. Study area | 8 |
| 2.1.1. Wild carnivore occurrence data | 8 |
| 2.1.2. Hunting yield data | 8 |
| 3. Methodology | 10 |
| 3.1. Models based on occurrence data | 10 |
| 3.2. Models based on hunting yield data | 10 |
| 3.3. Comparison of models based on hunting yield and occurrence data for red fox and badger | 11 |
| 4. Results and discussion | 11 |
| 4.1. Occurrence data models | 11 |
| 4.2. Models based on hunting yield data | 16 |
| 4.3. Comparison of models based on hunting yield and occurrence data | 18 |
| 4.4. Conclusions and further steps | 19 |
| References | 21 |
| Glossary | 25 |

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) were (deliverable 4.2) to develop a distribution/abundance model for carnivores at European scale, which is due on December 2022.

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.

Moreover, the *ENETWILD* consortium developed a generic model framework for predicting habitat suitability from occurrence data and to model HY densities first for wild boar, as a first study case, and afterwards for widely and constrained distributed ruminant species across Europe. Overall, the model frameworks developed well for these ungulates' species (<https://efsa.onlinelibrary.wiley.com/doi/10.2903/sp.efsa.2022.EN-7631>).

In parallel, during the last year *ENETWILD* consortium has also done an effort to collect information regarding the geographical distribution and hunting yield data (HY) of carnivores over Europe at different spatial resolution (e.g., hunting ground, municipalities, NUT3).

The general goal of this report is to apply the models developed for ungulates to carnivore's species and specifically to explore if these approaches can be used (i) for modelling the occurrence at European scale, (ii) implementing hunting yield (HY) density model framework for red fox and badger, and (iii) comparing the output of occurrence with observed HY.

1.2. Environmental variables and other predictors

Following to previous reports (e.g., *ENETWILD* consortium et al., 2021) we selected environmental variables closely related to wildlife 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.

| 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 reports (e.g. 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.

2.1.1. Wild carnivore occurrence data

Sightings records (presences) from the past 30 years (1990-2020) were obtained from collections hosted on the Global Biodiversity Information Facility (GBIF; extracted on 04/11/2022 <https://doi.org/10.15468/dl.c7v9a6>) which included records from our own *MammalNet* data collection project (*iMammalia and Mammalweb*; ENETWILD consortium et al., 2018). 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 as described by ENETWILD consortium et al., (2022b); to create, firstly, a presence-only dataset to supplement expert drawn estimates of species range, and then, 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 also ENETWILD consortium et al., 2022a for more details).

2.1.2. Hunting yield data

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 (Figure 1). It must be noted that the allowance of hunting both species is different within European countries. Hunting red fox is allowed for all countries, while hunting badgers is forbidden in 10 of the 33 UE countries (Kurek et al. 2020), thus almost one third of the study region could not report information about hunting statistics.

In line with previous reports, we have focused on maximum annual HYs between 2012-2021 hunting seasons and prioritized highest spatial resolution data. Due to the later, data selected for Sweden corresponded to 2019 and 2020 hunting season as it had the best spatial resolution (municipality level) than previous years (NUTS 3). Data collected for Poland 2021 and Hungary 2014 to 2016 were discarded, as they had management units (hunting grounds) discordant with other years' databases. Thus, they could not be included for calculating the maximum HY.

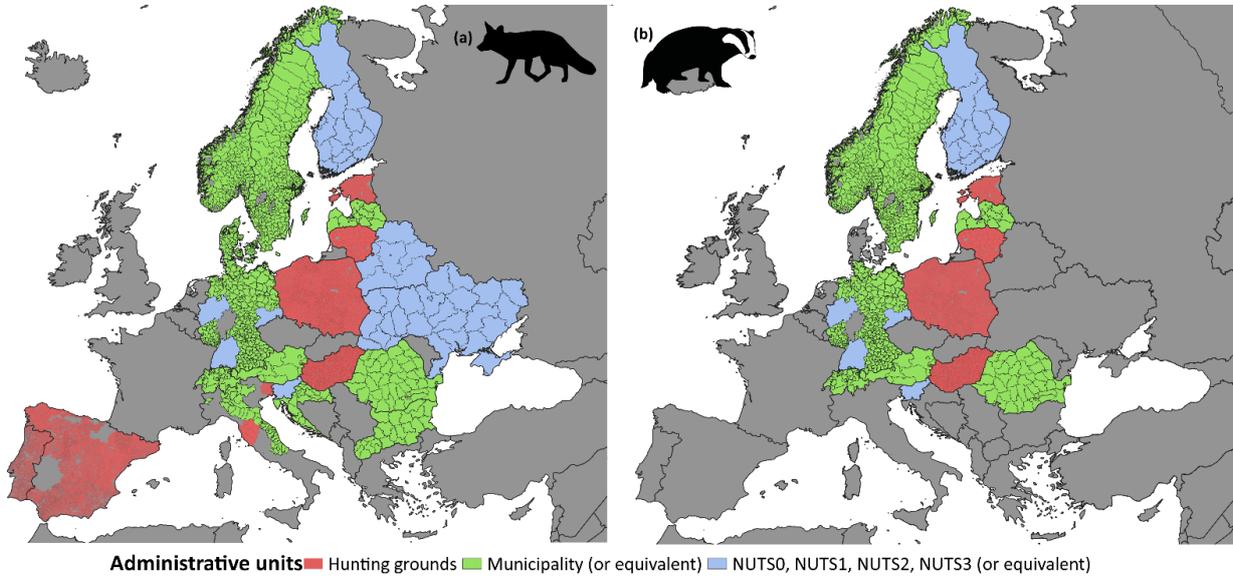


Figure 1³: Maps showing the extent of the study area. Different colours show the different administrative spatial data resolution received (from lowest level 'hunting ground' to highest level 'NUTS 0, NUTS 1, NUTS 2 or NUTS 3') in the models for (a) red fox and (b) badger.

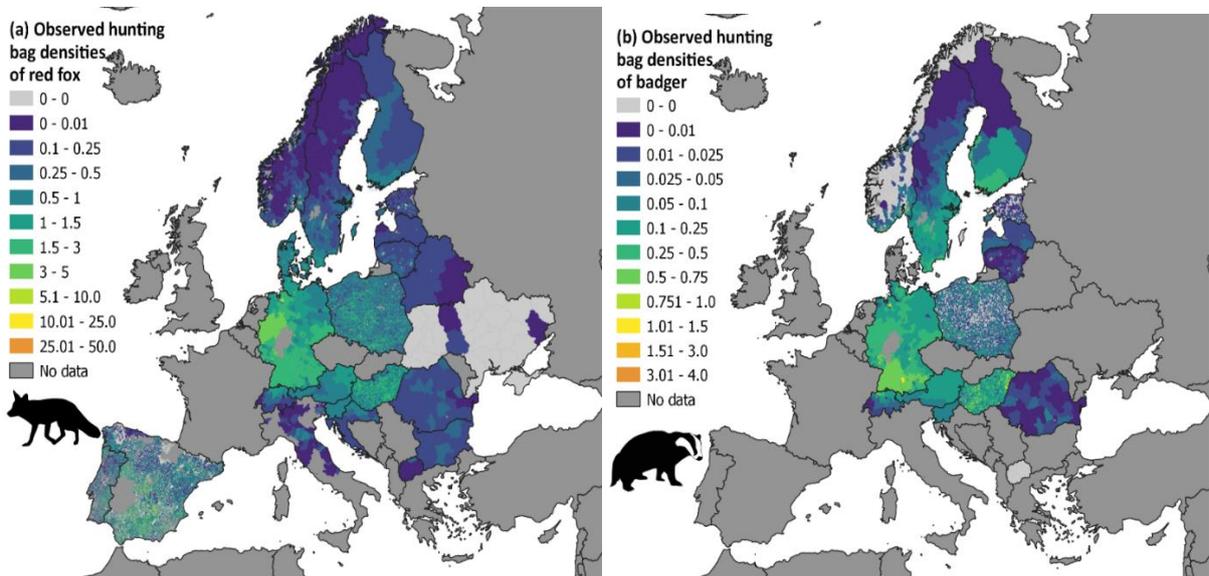


Figure 2⁴: Maximum hunting yield density data (2012-2021 hunting seasons) used for parameterizing (a) red fox and (b) badger species models according to territorial units (observed hunting bag density per km²).

³ **Map disclaimer:** The designations employed and the presentation of material on any maps included in this scientific output do not imply the expression of any opinion whatsoever on the part of the European Food Safety Authority concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

⁴ **Map disclaimer:** The designations employed and the presentation of material on any maps included in this scientific output do not imply the expression of any opinion whatsoever on the part of the European Food Safety Authority concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

3. Methodology

3.1. Models based on occurrence data

Following the existing methodology described by ENETWILD consortium et al. (2022b), we first estimated a broad-scale “stable” range for each species combining expert-derived maps with the presence-only records we had extracted from GBIF. We used these ranges to filter our binomial dataset to ensure a condition of stationarity was observed before using the resulting dataset to produce a finer-scale (2x2 km resolution) description of occurrence. For this we used a classical stepwise approach, estimating locations of likely absence based on survey effort followed by fitting using a random forest model considering variables describing climate, land cover, topography, and human disturbance (Table 2). As we have previously, highly correlated variables were identified and excluded prior to fitting.

For each species we used models to project likelihood of occurrence (suitability) across the full extent of interest (Figure 3). Model performance was assessed using 4-fold cross validation to compute several common metrics for predictive accuracy aggregated (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). Model residuals were tested for evidence of spatial autocorrelation, which can confound inference, using mantel correlograms (Legendre & Fortin, 1989). Finally, to evaluate the transferability of our model predictions we performed 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 carnivores were hunting yield densities across Europe (maximum number of individuals annually hunted within 2012-2021 hunting seasons, divided by unit area in km²; hereafter HY; Figure 2). Hunting yield density records were transformed to density data multiplying their values by 10,000 for modelling purposes (to have integer response variable for the negative binomial models). One model per species was calibrated following two approaches depending on hunting restriction of species.

- Red fox is allowed to be hunted in all the EU, for that reason we followed the generic framework developed by ENETWILD consortium et al. (2022a) for widespread ungulates, in which 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) and zero values were removed from the response variable.
- Badger is forbidden to be hunted in many European countries (Kurek et al. 2022), so that and it was similar to constrained distributed ruminants, zero values of the response variable remained instead of being removed, as they must be related to important characteristics of the environment for low HY. Moreover, since in the countries that permitted hunting, it was thought that neither the type of hunting nor the hunting effort would be the same, it was developed a negative binomial mixed model using country as random effect factor was used (Mollie et al. 2017).

The same eco-geographical predictors than occurrence models were used as fixed effects. The 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).

The 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) with the mean predicted HY in the same intervals, 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 to assess the potential effect of zeros in the parameterization of the models.

Model output was statistically downscaled to make predictions at 10x10km using EEA grid (<https://www.eea.europa.eu/data-and-maps/data/eea-reference-grids-2>). As countries were used as random effect for badger, the random effect of Austria was used for model projection as it had the closer value to the mean of the random estimates.

3.3. Comparison of models based on hunting yield and occurrence data for red fox and badger

To validate suitability of HY results, calibration plots were developed to assess the relationship between predicted suitability with observed HY of red fox and badger. 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

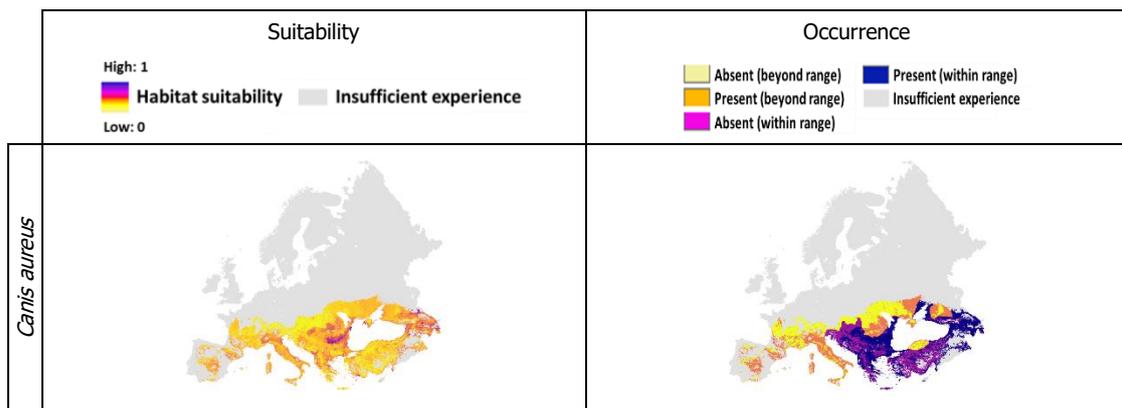
4.1. Occurrence data models

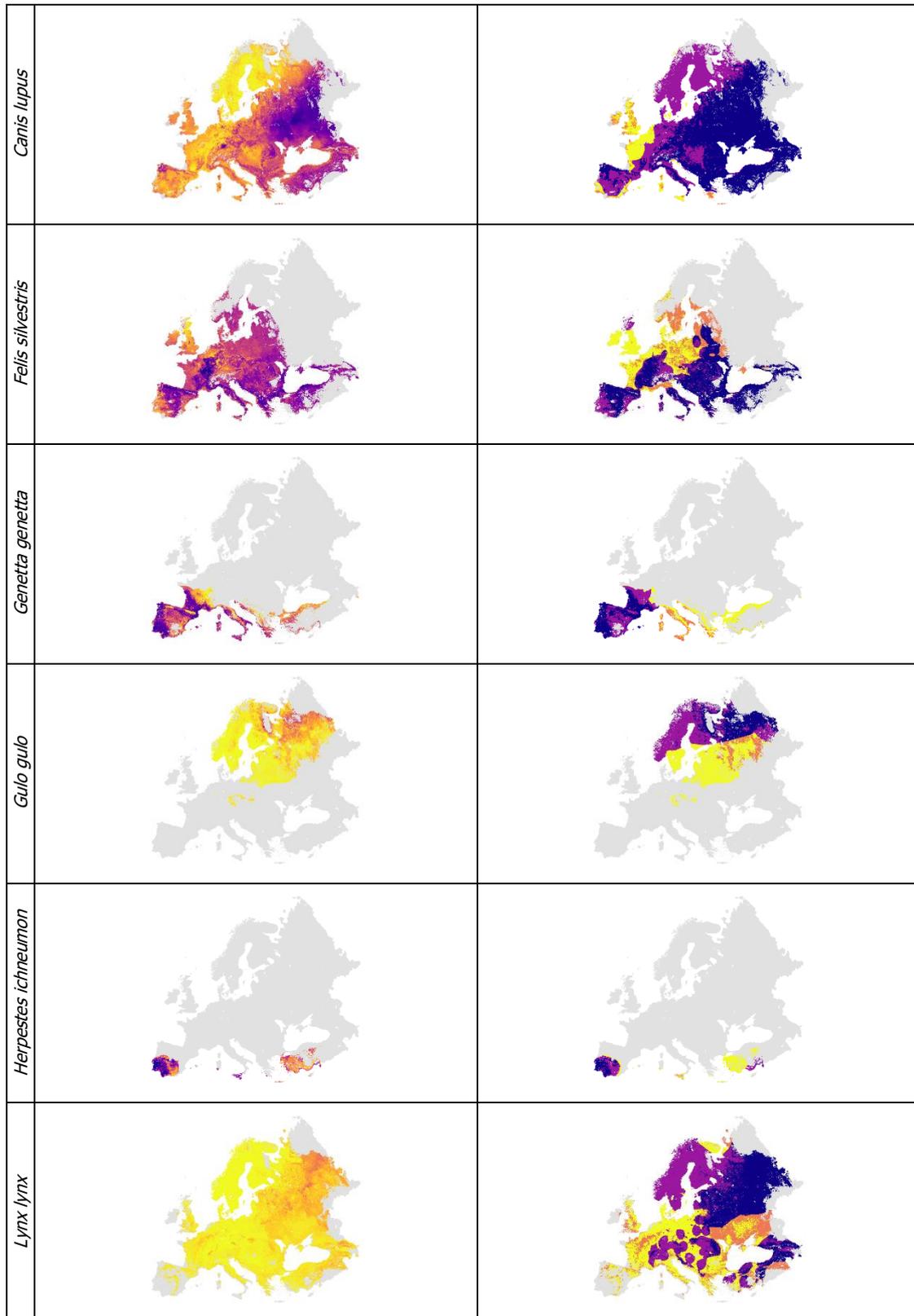
Sufficient data was available to make predictions for 21 of the 22 *Carnivora* species considered by this study (16 with corresponding model evaluation). Of those where prediction (and evaluation) was possible our results (Table 2) showed a continuation of the generally good evaluation statistics observed for ungulate species in our previous report (ENETWILD consortium et al., 2022b). The AUC statistic for all species except for three (*C. aureus*, *G. gulo* and *M. putorius*) exceeded 0.7 indicating good predictive performance reaching as high as 0.86 for *F. silvestris* and *M. meles*. Other threshold-dependent metrics (sensitivity, specificity, and True Skill Statistic) were also high with typically good balance between sensitivity and specificity (i.e., equally good prediction of absence as presence).

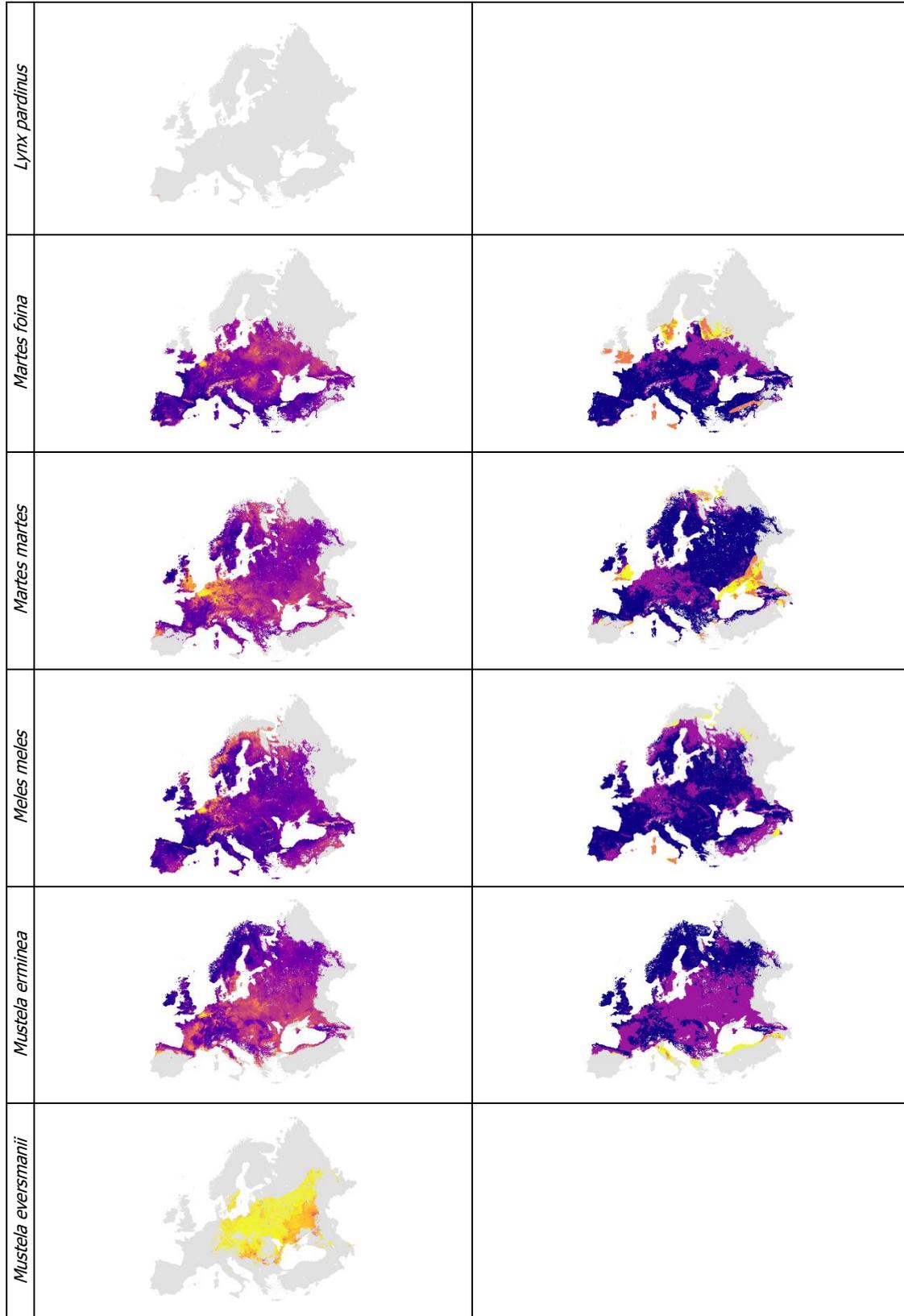
Predicted suitability, occurrence and transferability distributions are shown in Figure 3 for comparison with outputs of models based on hunting-yield. These maps appear mostly realistic for the distribution, although some species, such as *C. aureus* are known to be expanding, so the predicted distribution is not an accurate reflection of its current distribution but is a good reflection of its historic range. Whereas for invasive non-native species, such as *N. procyonoides* and *P. lotor*, and species with a very limited distribution within Europe (e.g., *L. pardinus* and *M. eversmannii*) reliable maps could not be produced.

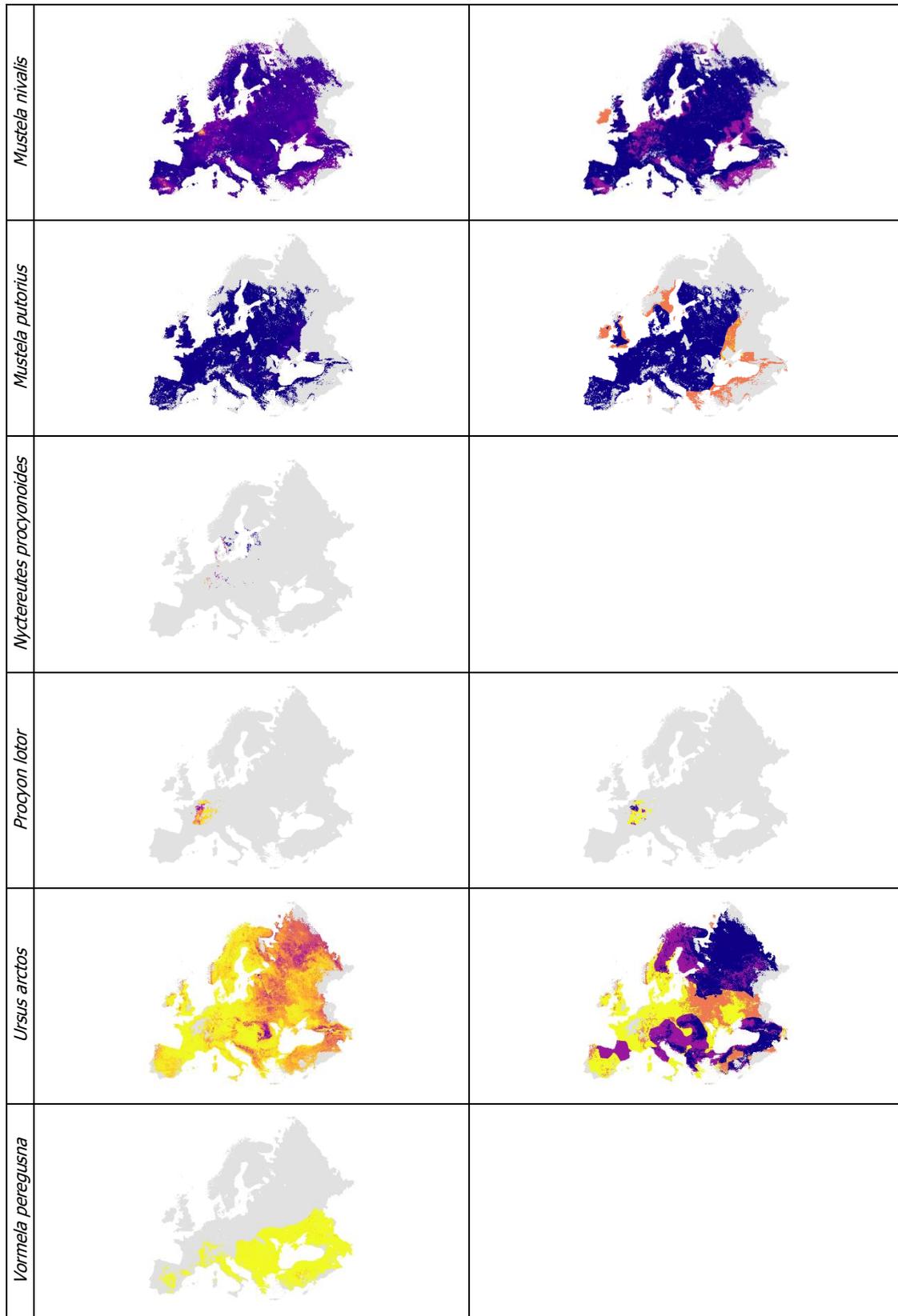
Table 2. Summary of distribution statistics; estimated “stable” range size (Rng), occupancy within range (Occ) and potential occupancy (Pot) across the model extent derived from thresholded model outputs, Figure 3, occurrence, including areas where prediction may be uncertain, in km²), predicted observability (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 | Rng | Occ. | Pot. | Obs. | AUC | THD | SE | SP | TSS |
|---------------------------------|------------|-----------|-----------|------|------|------|------|------|------|
| <i>Canis aureus</i> | 3,651,736 | 1,172,856 | 4,685,148 | 0.72 | 0.64 | 0.13 | 0.87 | 0.49 | 0.36 |
| <i>Canis lupus</i> | 19,038,912 | 6,286,092 | 6,544,984 | 0.08 | 0.81 | 0.28 | 0.68 | 0.85 | 0.53 |
| <i>Felis silvestris</i> | 8,134,280 | 2,991,824 | 7,387,324 | 0.04 | 0.86 | 0.54 | 0.80 | 0.80 | 0.59 |
| <i>Genetta genetta</i> | 1,841,376 | 466,232 | 681,508 | 0.14 | 0.84 | 0.54 | 0.75 | 0.78 | 0.53 |
| <i>Gulo gulo</i> | 5,008,648 | 1,326,148 | 1,913,940 | 0.29 | 0.62 | 0.10 | 0.81 | 0.92 | 0.73 |
| <i>Herpestes europunctatus</i> | - | - | - | - | - | - | - | - | - |
| <i>Herpestes ichneumon</i> | 882,408 | 125,108 | 194,596 | 0.17 | 0.79 | 0.73 | 0.72 | 0.77 | 0.49 |
| <i>Lynx lynx</i> | 11,910,944 | 3,292,400 | 5,087,752 | 0.42 | 0.75 | 0.05 | 0.71 | 0.72 | 0.43 |
| <i>Lynx pardinus</i> | - | - | - | - | - | - | - | - | - |
| <i>Martes foina</i> | 11,940,288 | 3,620,132 | 4,312,300 | 0.06 | 0.84 | 0.65 | 0.75 | 0.77 | 0.51 |
| <i>Martes martes</i> | 17,180,976 | 6,307,468 | 7,257,956 | 0.07 | 0.86 | 0.58 | 0.80 | 0.79 | 0.59 |
| <i>Meles meles</i> | 16,936,992 | 5,246,252 | 5,428,584 | 0.08 | 0.85 | 0.72 | 0.79 | 0.74 | 0.53 |
| <i>Mustela erminea</i> | 17,629,592 | 3,992,132 | 4,029,992 | 0.04 | 0.85 | 0.69 | 0.79 | 0.76 | 0.55 |
| <i>Mustela eversmanii</i> | - | - | - | - | - | - | - | - | - |
| <i>Mustela nivalis</i> | 20,612,864 | 7,738,672 | 7,823,488 | 0.02 | 0.80 | 0.84 | 0.75 | 0.73 | 0.47 |
| <i>Mustela putorius</i> | 12,236,808 | 5,984,684 | 9,228,764 | 0.01 | 0.65 | 0.93 | 1.00 | 0.40 | 0.40 |
| <i>Nyctereutes procyonoides</i> | - | - | - | - | - | - | - | - | - |
| <i>Procyon lotor</i> | 98,728 | 18,120 | 876,672 | 0.17 | 0.79 | 0.44 | 0.95 | 0.80 | 0.75 |
| <i>Ursus arctos</i> | 11,951,144 | 3,747,288 | 5,341,148 | 0.55 | 0.85 | 0.11 | 0.85 | 0.78 | 0.63 |
| <i>Vormela peregusna</i> | - | - | - | - | - | - | - | - | - |
| <i>Vulpes lagopus</i> | - | - | - | - | - | - | - | - | - |
| <i>Vulpes vulpes</i> | 20,757,904 | 9,185,492 | 9,194,104 | 0.16 | 0.81 | 0.77 | 0.82 | 0.65 | 0.48 |









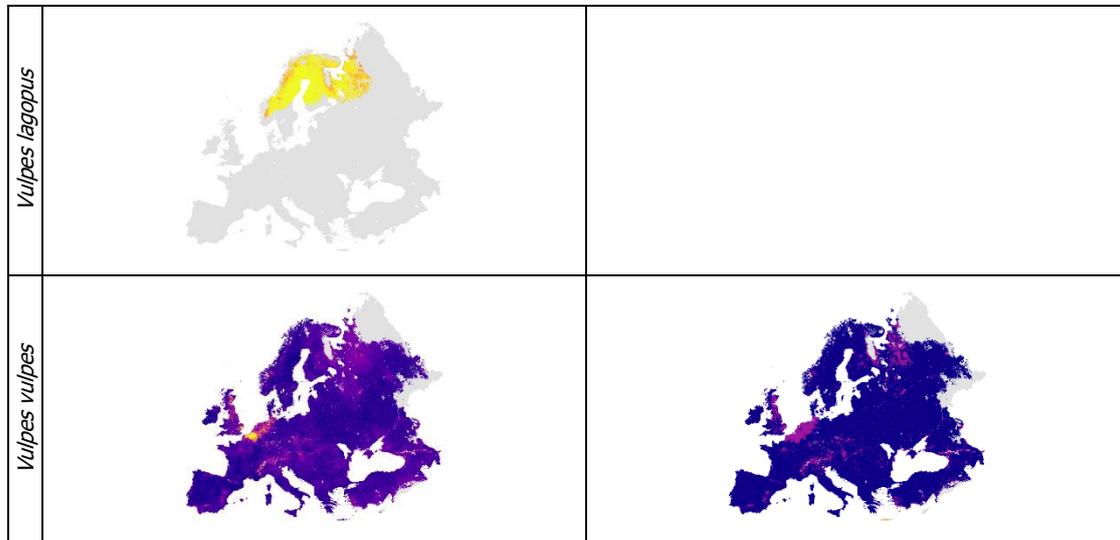


Figure 3⁵: Predictions of carnivore occurrence data models. Mapped output showing projected habitat suitability (mean across repetitions ignoring individual exclusions where model transferability may be limited) and threshold occurrence (based on mean habitat suitability and estimated threshold using 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.

4.2. Models based on hunting yield data

The calibration plot of red fox HY showed an irregular predictive performance for the bins, being good for the first classes, but overestimating prediction for middle and last classes, indicating that middle to high density predictions could be improved (Figure 4a).

On the contrary, badger calibration plot showed a better predictive performance than the red fox. All the classes except the last bin, which showed overestimated prediction, had a good predictive performance. So it was considered that a general good accuracy of the badger density predictions was achieved, except for the highest density prediction for which accuracy could be improved (Figure 4b).

The division of the calibration plot per each bioregion for red fox showed a good performance for the Eastern and Northern bioregion, an overfitted performance for the Southern bioregion and a good performance for first bins of Western bioregion, but the pattern become overfitted from medium to high density predictions (Figure 4c). It happened similarly with the division of the calibration plot per bioregion for the badger, which showed also a general good performance for the Eastern and Northern bioregion, but a changing pattern for the Western bioregion showing under or overfitting pattern depending on the bin (Figure 4d).

⁵ **Map disclaimer:** The designations employed and the presentation of material on any maps included in this scientific output do not imply the expression of any opinion whatsoever on the part of the European Food Safety Authority concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

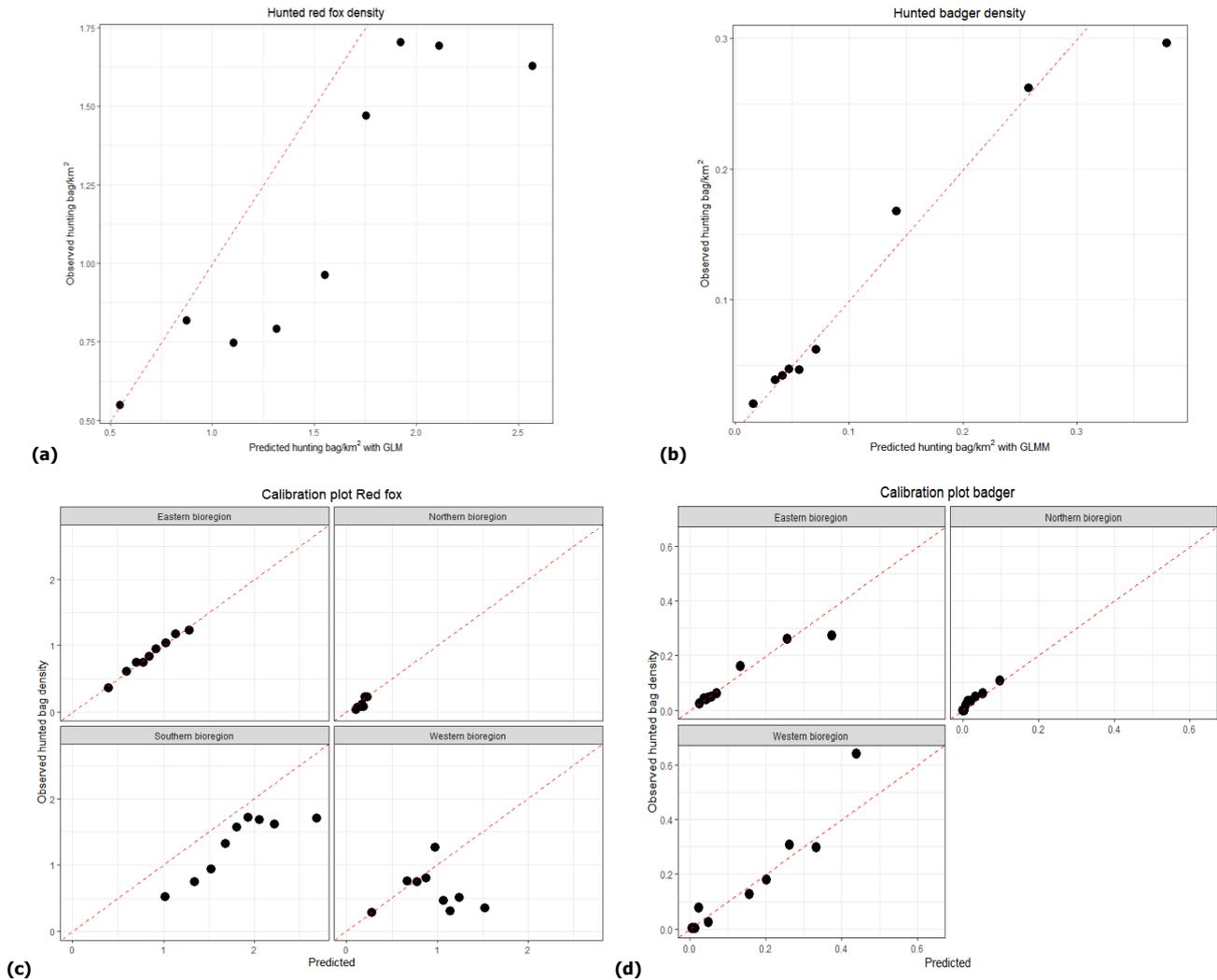


Figure 4: Calibration plot for assessing predictive performance of (a) red fox HY model and (b) badger HY model for all Europe and per bioregions (c) for red fox and (d) for badger. Plots show the relationship between the predicted hunting yield densities (HY) and the observed ones on the validation datasets.

HY models’ predictions suggested different patterns depending on the species (Figure 5).

The highest HY predictions for red fox were reached in the Iberian Peninsula (Figure 5a). Although, the abundance distribution pattern was widely scattered over all Europe with medium values of HY predictions. This was expected for a widely distributed species which reflected its high ecological plasticity. Lower HY density values were found in the Nordic countries, although nowadays red fox seems to be expanding its range in these countries and displacing the Arctic fox (Rød-Eriksen et al. 2019), so HY densities in the following years would be expected to increase in the Nordic countries. To sum up it seemed that red fox had a HY density gradient pattern, reaching higher densities at Southern latitudes and decreasing its HY density towards Northern latitudes and higher altitudes.

On the other hand, although the badger can also be considered a widely distributed species in Europe, hunting status changed within European countries, being a game species in some countries while in others it is protected (Kurek et al., 2020). Concerning badger HY density model, according to model projections medium density values would be found overall in Europe, reaching its highest densities at South of Romania, and lowest densities in mountain ranges and Nordic

countries (Figure 5b). Badger projections were limited to the regions where data were collected, as there were limitation due to environmental variables according to MESS analysis.

Both species are widely distributed, according to HY density results the red fox had higher densities than badgers in Europe (Figure 5a and 5b). Moreover, comparing both models, it seemed that in some countries, such as Germany, Austria, Czech Republic, Poland, higher HY densities of red fox overlap with a low-density pattern of badgers and *vice versa*. This observed pattern could be revealing two different reasons. In the first place, it could be due to a biological reason, and that spatial interspecific interaction of both species would take place only at medium densities of both species, while high densities of one species would provoke spatial avoidance of the other or different habitat selections (Trewby et al. 2008). In the second place, it may be not reveal a real spatial interspecific avoidance between species, but a difference on hunting tradition of the species in these places, showing a possible preference of hunters for hunting one or another species. On the contrary, other countries such as Switzerland or Austria showed similar density pattern of both species, what could be indicating that hunter's preference was similar for both species (Figure 2).

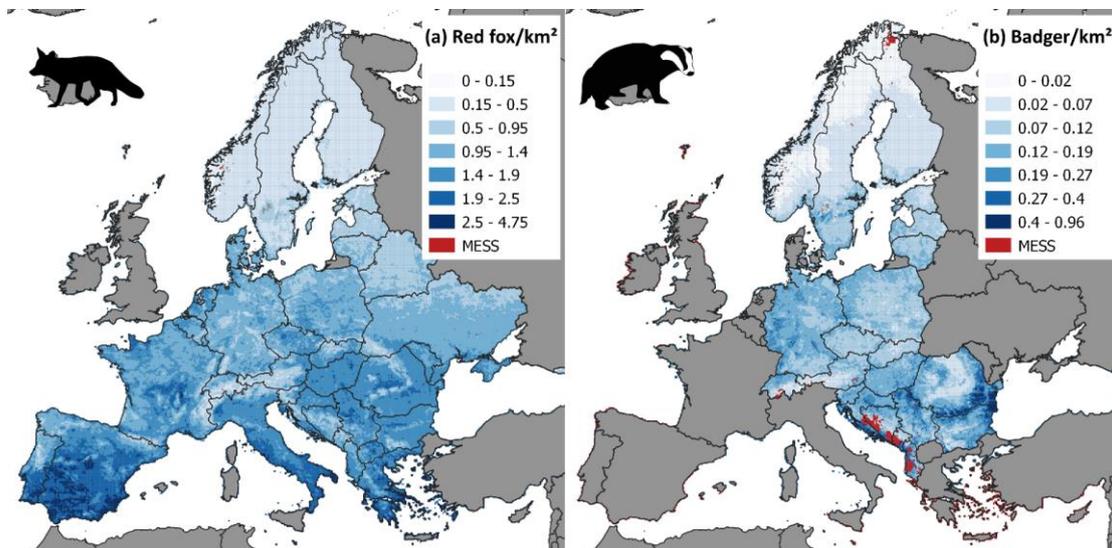


Figure 5⁶: Hunting yield (HY) density (individual hunted per km²) of red fox (a), and badger (b) at 10x10km. Red areas are beyond the environmental domain according to MESS analyses.

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

The observed HY and habitat suitability showed a flat pattern for red fox remaining constant across habitat suitability values if it is considered the complete dataset, not being possible to use suitability as a proxy of observed HY. However, if zero density values are removed from the dataset, it can be distinguished two intervals for determining low, and medium-high values of observed HY from suitability (Figure 10a).

Contrary to the above red fox pattern, the one for badger since the complete dataset showed a linear pattern in which all the intervals can be distinguished, being possible to use suitability as a

⁶ **Map disclaimer:** The designations employed and the presentation of material on any maps included in this scientific output do not imply the expression of any opinion whatsoever on the part of the European Food Safety Authority concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

proxy of observed HY. It could be clearly determined three intervals low, medium and high HY from suitability values. However, if zero values are removed from the dataset the pattern become flatted and suitability values would not be used as a proxy of observed HY (Figure 10b).

Finally, it must be added that the results obtained for the relationships between observed HY and suitability were in agreement. Although only the final HY models have been presented in this document, different models were tested using either the full database, or alternatively by assessing the removal of zero values. Consistent with these results (Fig. 6), the best performance was obtained when considering the non-zero data for the red fox, while for the badger the best model was generated using the full database including zeros (see methodological framework 3.2).

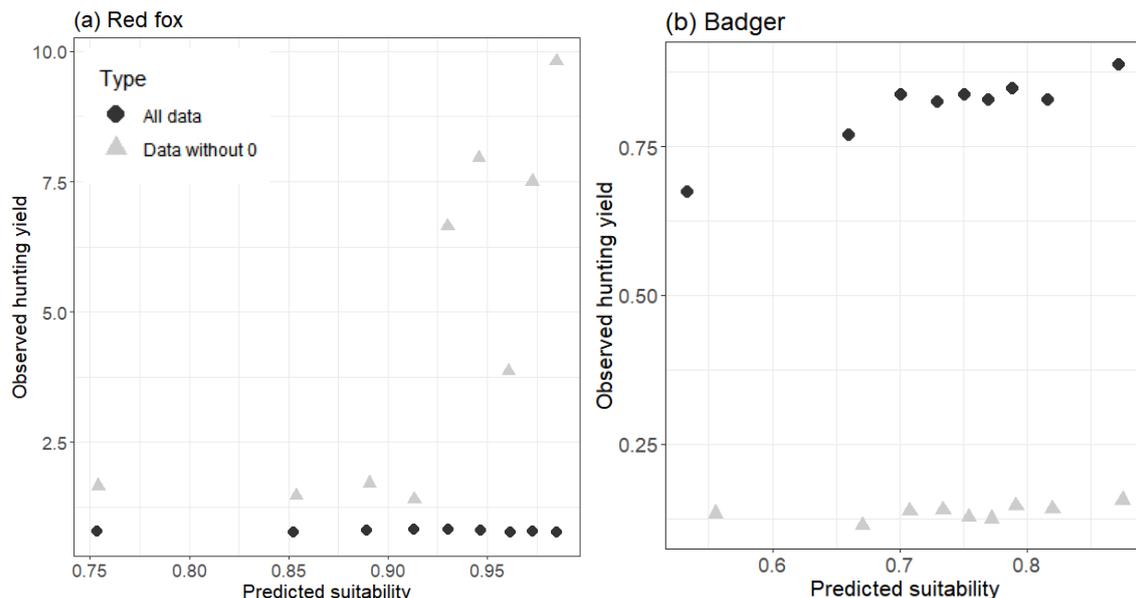


Figure 6: Calibration plots of (a) red fox and (b) badger showing the relationship between predicted suitability and observed hunting yields values.

4.4. Conclusions and further steps

Occurrence data model

- We have previously highlighted the potential use of variable importance as an additional tool for model validation alongside statistical metrics of performance. While such information is simple to extract from models it's interpretation can be more difficult and further work is ongoing to ensure any inference is robust.
- As part of the modelling approach, for simplification, we assume that observability is constant across our extent of interest. However, several factors may confound this assumption: substantial variation in species richness / behaviour (ecological) and recorder behaviour (anthropological). Additional work is required to assess the extent of any variation and the impact of any bias it may introduce in model predictions.

Hunting yield density data model

- The difference in hunting traditions and conservation status within each country on carnivore species must be considered for modelling carnivore HY.
- The framework proposed for widely distributed ungulates seemed to be a good approximation and consistent for modelling widely distributed and hunted carnivores' species such as the red fox, although results showed a slight overprediction.

- A new framework proposed for the badger seemed to be a good approximation for modelling the abundance distribution of species that may have different legislation and hunting tradition between countries.
- Regarding the results obtained with the new framework developed for badgers, the red fox models could consider implementing this framework and evaluate if results improve in comparison with the widely distributed ungulates framework.
- Model projections showed, in general, good abundance patterns for red fox and badger. There was not a single framework for modelling carnivores at the European scale and it should be adapted to the particularities of the distribution of the dataset for modelling. Moreover, using or removing zeros from the datasets for HY modelling was important and should be considered when modelling.
- It should be explored how to improve the results and projection of the HY models when the unavailability of hunting activity for some species limits the extrapolation to other regions.

Validation of suitability on HY

- Suitability has potential to be used as a proxy for abundance of red fox and badger HY. As it happened for ungulate species, the resolution of intervals for using suitability as a proxy of abundance will depend on each 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. 2022a. Update of model for wild ruminant abundance based on occurrence and first models based on hunting yield at European scale. EFSA Supporting Publication, 19(2), EN-7174. 30 pp.
- ENETWILD-consortium, Illanas, S, Croft, S, Smith, GC, López-Padilla, S, Vicente, J, Blanco-Aguilar, JA, Scandura, M, Apollonio, M, Ferroglio, E, Zanet, S, Vada, R, Keuling, O, Plis, K, Podgorski, T, Brivio, F, Fernández-López, J, Ruiz, C, Soriguer, RC, Acevedo, P. 2022b. New models for wild ungulates occurrence and hunting yield abundance at European scale. EFSA Supporting Publications, 19(10), EN-7631. 43 pp.
- Rød-Eriksen L, Skrutvold J, Herfindal I, Jensen H, Eide N.E. 2019. Highways associated with expansion of boreal scavengers into the alpine tundra of Fennoscandia. *J Appl Ecol.* 57: 1861– 1870. <https://doi.org/10.1111/1365-2664.13668>
- 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.

- Kurek P, Piechnik Ł, Wiatrowska B, Ważna A, Nowakowski K, Pardavila X, Cichoński J, Seget B. Badger *Meles meles* as Ecosystem Engineer and Its Legal Status in Europe. *Animals* 2022, 12, 898.
- 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, Pokorný 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.
- Mollie E. Brooks, Kasper Kristensen, Koen J. van Benthem, Arni Magnusson, Casper W. Berg, Anders Nielsen, Hans J. Skaug, Martin Maechler and Benjamin M. Bolker (2017). *glmmTMB* Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal*, 9(2), 378-400. doi: 10.32614/RJ-2017-066.
- 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.
- 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.
- Trewby ID, Wilson GJ, Delahay RJ, Walker N, Young R, Davison J, Cheeseman C, Robertson PA, Gorman ML, McDonald RA. 2008. Experimental evidence of competitive release in sympatric carnivores. *Biology Letters* 4:170-172.
- 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 Edicions. 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. |
| Bayesian hierarchical framework | Statistical model written in multiple levels (hierarchical form) that estimates the parameters of the posterior distribution using the Bayesian method |
| 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 | Datasets 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. |
| Random forest | Modelling algorithm based on regression trees. |
| 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. |