



EXTERNAL SCIENTIFIC REPORT

APPROVED: 09 August 2021

doi:10.2903/sp.efsa.2021.EN-6825

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

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

Abstract

In the previous ENETWILD model, the predicted patterns of wild boar abundance based on hunting yield data reached an acceptable reliability when the model was downscaled to higher spatial resolution. This new approach, based on the modelling of hunting yield densities instead of hunting yield counts and the assessment of spatial autocorrelation, was only applied with simulated data and with data from two regions at hunting ground level, the smallest spatial resolution. In this report, (1) we evaluate whether this approach can correct the overpredictions for high-resolution predicted patterns when raw data are present at a different spatial resolution (*i.e.* the European region). For this purpose, hunting yield densities were incorporated as response variable (one model per bioregion) and predictions reliability at 10x10km and 2x2km spatial resolution were assessed. Internal validations and comparisons with the previous two-step model carried out at European scale were addressed, as well as an evaluation with external data at the same scale at country level. The model presented certain overprediction (much less than the previous model) of the total hunting bags reported per country, although a good correlation in terms of values and linearity between observed and predicted values was achieved. Secondly (2), a generic model framework to predict habitat suitability and likely occurrence for wildlife species using opportunistic presence data was proposed (occurrence records for wild ungulate species from the past 20 years exclusively from the Global Biodiversity Information Facility extracted on 9/12/2020). Across all wild ungulate species (elk (*Alces alces*), roe deer (*Capreolus capreolus*), red deer (*Cervus elaphus*), dam deer (*Dama dama*), muntjac (*Muntiacus reevesi*), wild boar (*Sus scrofa*)) the model framework performs well. For those species where area under the curve is below 0.7 we note lower accuracy in predicting absences, which requires further investigation to understand the root cause; whether a result of underlying assumptions regarding the testing data or due to the model performance itself.

© European Food Safety Authority, 2021

Key words: distribution, hunting bags, occurrence, population abundance, population monitoring, risk assessment, spatial modelling, wild ruminants, *Sus scrofa*, wild boar

Question number: EFSA-Q-2020-00678

Correspondence: alpha@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 ALPHA and DATA units, and ENETWILD partners for reviewing this manuscript. We are grateful to ENETWILD collaborators² and data providers and institutions feeding GBIF³ for sharing data on wild boar and other species.

Suggested citation: ENETWILD consortium, Illanas S, Croft S, Smith G C, Fernández-López J, Vicente J, Blanco-Aguiar J A, Pascual-Rico R, Scandura M, Apollonio M, Ferroglio E, Keuling O, Zanet S, Brivio F, Podgorski T, Plis K, Soriguer R C and 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. 30 pp. doi:10.2903/sp.efsa.2021.EN-6825

ISSN: 2397-8325

© European Food Safety Authority, 2021

² <http://www.enetwild.com/collaborators>

³ <https://doi.org/10.15468/dl.v2ydsu>

Summary

Background and objectives: In May 2020 the ENETWILD consortium updated the relative abundance maps based on hunting yield (HY) data at 10x10km resolution (doi: 10.5281/zenodo.3702131, <https://zenodo.org/record/3702131#.YPbQcej7TD5>). The best updated model followed a two-step independent bioregion modelling approach that allowed for flexibility in predictor effects avoiding abrupt changes in predictions across bioregion boundaries (for modelling purposes Europe is subdivided in four bioregions, in order to subgroup the data). While this approach showed consistency with the spatial pattern shown in previous reports, hunting yield model predictions overestimated the hunting bag numbers obtained from the external datasets. For this reason, in November 2020 new approaches were developed, based on the modelling of HY densities instead of HY counts and the assessment of its spatial autocorrelation. These approaches were applied to simulated data and with data from two regions at hunting ground level resolution (*i.e.* the smallest spatial resolution available).

When models are parameterized to broad territories (*i.e.* the European region), available raw data is normally collected at different spatial resolution (e.g. hunting ground, municipalities, NUT3), as it is our case in previous and in the present report. The goal of this report is to evaluate whether the new approach is capable to correct the overpredictions for high-resolution predicted patterns when raw data are collected at different spatial resolution. Therefore, hunting yield densities (HY) as response variable (one model per bioregion) were incorporated and predictions reliability at 10x10km and 2x2km spatial resolution were assessed.

In addition, a generic framework is proposed to predict habitat suitability and likely occurrence for large terrestrial mammal species using opportunistic sightings data was proposed. This report presents first spatial distribution models developed by ENETWILD based on occurrence data for six wild ungulate species (elk (*Alces alces*), roe deer (*Capreolus capreolus*), red deer (*Cervus elaphus*), dam deer (*Dama dama*), muntjac (*Muntiacus reevesi*), wild boar (*Sus scrofa*)).

Data: For this report, we used HY data compiled for the period 2014-2019 from records submitted to the ENETWILD Wild Boar Data Model (WBDM), extracted on 31/07/2020, and occurrences available from the past 20 years (since 2020) through the Global Biodiversity Information Facility (GBIF) for suitability models, extracted on 9/12/2020⁴.

Modelling: Similar to models based on HY counts, the response variable was the maximum number of wild boar annually hunted in 2014-2018 hunting seasons divided by the area (km²) of the corresponding administrative unit (HY densities). To avoid abrupt transitions among neighbouring bioregions and allow more flexibility in predictor responses, four models (one per each bioregion) were developed.

Results: Predictions for the HY density model presented similar pattern to the two-step modelling approach, where the highest predicted abundances were reached at central Europe. The external evaluation of the HY models presented certain over-prediction of the total hunting bags reported per country when it was downscaled to 10x10 km grid. However, there was a significant reduction of wild boar HY amount predicted with respect to the previous two-step model, and better correlation between observed and predicted values (in terms of values and linearity). The HY models were also downscaled to 2x2km grid, and the predictive performance was similar to that obtained at 10x10km grid.

For occurrence-based models, new insights into the presence and absence of species within their current range as well as their potential for inhabiting new regions across Europe were provided. Overall, the model framework performs well with good predictive performance for all species. For those species where area under the curve (AUC) is below 0.7 we note lower accuracy in predicting

⁴ <https://doi.org/10.15468/dl.xa9asm>

absences, which requires further investigation to understand the root cause; whether a result of underlying assumptions regarding the testing data or due to the model performance itself.

Conclusions: in relation to the modelling approaches, it can be concluded that:

- By modelling density of hunted wild boar rather than hunting yields counts, models are more reliable to predict the spatial pattern of wild boar abundance at European scale.
- Including the area of the territorial units within the response variable (by modelling density of hunted wild boar) the parameterized models are more flexible to transfer among spatial scales, and overprediction associated to downscaling in previous report is here mostly avoided.

Concerning the generic suitability model the following can be concluded:

- Overall, initial results of habitat suitability models are encouraging suggesting that a relatively generic framework of this type could be successfully applied to inform occurrence across different wild ungulate species of interest because it meets accuracy and readiness.
- Within the constraints of the available data for habitat suitability models (probability of presence) our modelling approach appears to perform reasonably well; area under the curve (AUC) generally close to or above 0.7 (note that estimates are not inflated by spatial autocorrelation which is often the case in published studies and that absences are not certain).
 - Current modelling choices (i.e. modelling of HY densities instead of HY counts to correct the overpredictions for high-resolution predicted patterns when raw data are collected at different spatial resolution) do potentially reduce the usefulness of model output specifically regarding the transferability of model projections.
 - Limiting model fitting to regions where species are considered "stable" as defined by the International Union for Conservation of Nature (IUCN, for many species including European rabbit for example published ranges represent the extent of native populations which are inherently more restrictive than their naturalised "stable" range severely impacting model transferability).

Next steps for data acquisition and modelling are:

- Data about wild boar abundance at the highest resolution for each country are scarce and patchy. Therefore, a complete compilation of data for different spatial resolution should be carried out for modelling temporal dimension of wild boar abundance patterns.
- Temporality in the wild boar abundance model parameterization and predictions should be explored.
- In the previous report with simulations and regional data we showed that accounting for spatial autocorrelation the predictive performance of the models is improved. Feasible approaches to manage spatial autocorrelation at European scale should be assessed.
- There are several alternative approaches and improvements for habitat suitability modelling (probability of presence) which could be adopted to mitigate this issue in the general framework, applying a more data-driven approach to geographically delimit model extents. Additional fine tuning is required to address several outstanding issues before future updates. In particular, the development of a data-driven approach to geographically delimit model extents, excluding regions where species distribution may be non-stationary (range edges), which will hopefully maximise model transferability. Greater flexibility for species specificity in the model framework will likely also need to be considered and balanced against the efficiencies of the generic approach to drive

continued improvement in model performance. It will be important to balance the transparency and speed of the generic approach against more flexible (and often argued more accurate but likely slower) species specific adaptation.

Table of contents

Abstract	1
Summary	3
1. Introduction	7
1.1. Background and Terms of Reference as provided by the requestor	7
1.2. Scope of the report	7
1.3. Environmental variables and other predictors	7
2. Data	9
2.1. Study area	9
2.2. Data collection	10
2.2.1. Wild boar hunting bag data	10
2.2.2. Wild ungulates occurrence data	12
3. Methodology	13
3.1. Wild boar: models based on hunting yield data	13
3.1.3. Model comparisons	15
3.2. Models based on occurrence data	15
3.2.1. Bias correction	15
3.2.2. Model fitting	16
3.2.3. Model evaluation	16
4. Results and discussion	17
4.1. Wild boar: Model based on hunting yield data	17
4.2. Wild ruminants and wild boar: occurrence data model	22
4.3. Conclusions and further steps	24
References	26
Glossary	28

1. Introduction

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

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

The terms of reference for the present report (specific contract 8) were to update the occurrence and hunting yield-based data models for wild boar and other ungulates at European scale.

1.2. Scope of the report

The ENETWILD consortium (www.enetwild.com) implemented an EFSA funded project whose main objective has been the collection of information regarding the geographical distribution and abundance of wild boar and other 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 May 2020 ENETWILD presented two models of wild boar hunting yield predictions, from regional and local hunting records, at 10x10 km grid (ENETWILD consortium et al. 2020a). In November 2020 a new procedure was successfully tested at regional scale to make more accurate predictions (ENETWILD consortium et al. 2020). The first goal of this report is to evaluate whether the new approach is capable to correct overpredictions for high-resolution predicted patterns when raw data present different spatial resolution. In addition, this document outlines our proposed approach for predicting wild ungulate species habitat suitability (probability of presence) using the large volumes of opportunistic occurrence data collected mostly from citizen science projects. Whilst abundant this type of data is known to contain several sources of bias which make working with it a challenge. Nevertheless, it remains the best, most comprehensive, data source available for large scale projects such as ENETWILD which aims to provide insight across continental or global extents. As such the methodology we outline here contains features designed to mitigate as far as possible biases in the data providing a pure understanding of species ecology and the environmental drivers that their persistence in the landscape. Initially we consider large terrestrial mammal species typically surveyed using visual observation only. We further restrict application to a limited number of key species of interest (*Artiodactyla*) with native/established "stable" populations in Europe according to published IUCN ranges.

1.3. Environmental variables and other predictors

According to previous reports (ENETWILD consortium et al. 2018, 2019b, 2020c) 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.appolutelydigital.com/DataPrimer/part154.html>). The percentage of land that is suitable for wild boar in each administrative unit was quantified to be used as a predictor in the hunting yield data model. For this purpose, we delineated the suitable area for wild boar by reclassifying land use classes based on specific wild boar habitat preferences (see Table 1) and suitability definitions as in Alexander et al. (2016). 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 wild boar abundance model we decided to consider as predictor the percentage of *Eucalyptus* spp. as dominant species obtained from Brus et al. (2011) (European Forest Institute <https://www.efi.int/knowledge/maps/treespecies>). Raster predictor layers and grid polygons were managed using QGIS 3.4 and *rgdal*/R packages (Bivand et al. 2006).

Table 1: Variables used to model (i) the spatial pattern of wild boar (*Sus scrofa*) abundance and (ii) wild ungulates distribution (*Alces alces*, *Capreolus capreolus*, *Cervus elaphus*, *Dama dama*, *Muntiacus reevesi*, *Sus scrofa*) 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
SUIT	Percentage of habitat suitable	lc_190	Urban areas
NUT	Administrative level	lc_200	Bare areas
ASF	Years since the first ASF report	lc_201	Consolidated bare areas
BIOREG	Environmental bioregions	lc_202	Unconsolidated bare areas
ALT	Mean altitude	lc_210	Water bodies
AREA	Area of sampling unit	lc_220	Permanent snow and ice
SUIT	Percentage of habitat suitable	Eu	Percentage of <i>Eucalyptus sp.</i>

2. Data

2.1. Study area

The study area is the same as the previous report. It includes all countries in mainland Europe with the Ural Mountains as the eastern limit (ENETWILD consortium et al. 2019a, Figure 1). Spans 11,019,700 km² (110,197 10x10 km and 2,787,877 2x2 km grid cells) and includes Mediterranean islands and the UK and Ireland. It comprises a broader area than those reported as the distribution of wild boar for Europe according to the IUCN (<https://www.iucnredlist.org/species/41775/10559847>).

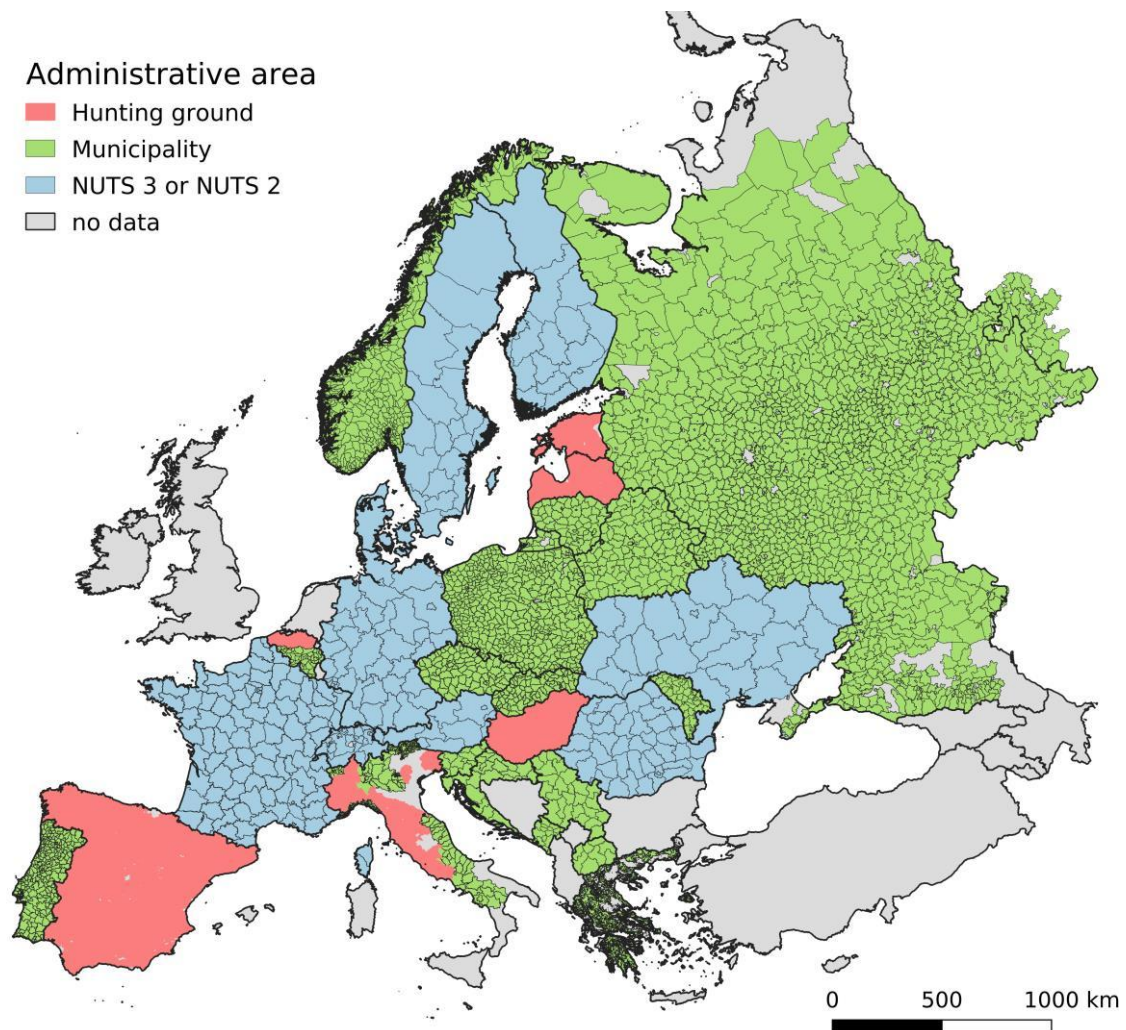


Figure 1: Map showing the extent of the study area. Different colours show the different administrative area levels used (from lowest level 'hunting ground' to highest level 'NUTS 3 or NUTS 2') in the models for wild boar density.

2.2. Data collection

2.2.1. Wild boar hunting bag data

For this report new hunting yield data was incorporated for modelling from the ENETWILD Wild Boar Data Collection Model (WBDM) downloaded 31/07/2020, specific dataset for, Czech Republic, Germany, Trento and Umbria (Italy), Kosovo, Lithuania, Poland, Slovenia and Castilla-La Mancha and Vizcaya (Spain) were added. We used 28,246 territorial units, of which 26,234 correspond to hunting grounds, 1,772 to municipality level (or equivalent) and 240 to NUT3 (or equivalent; see Figure 1).

As in previous reports, we focused on those units with at least one annual hunting bag record available (higher than 0) from 2014 to 2018 hunting seasons (Figure 2). Similarly, the maximum number of wild boar hunted per hunting season within the study period was considered for modelling but in this case the response variable was obtained by divided the number of hunted animals by area (km²), i.e., we modelled hunting yield (HY) density of wild boar. We also removed

density values above 50 wild boar hunted/km², as there are due to small areas which are topological errors from joining shapefile layers from different regions. Density data was transformed multiplying its values by 100 for modelling purposes (to have integer response variable for the negative binomial models). For modelling, data was separated by bioregion, resulting in four datasets (see Figure 3) and therefore four models.

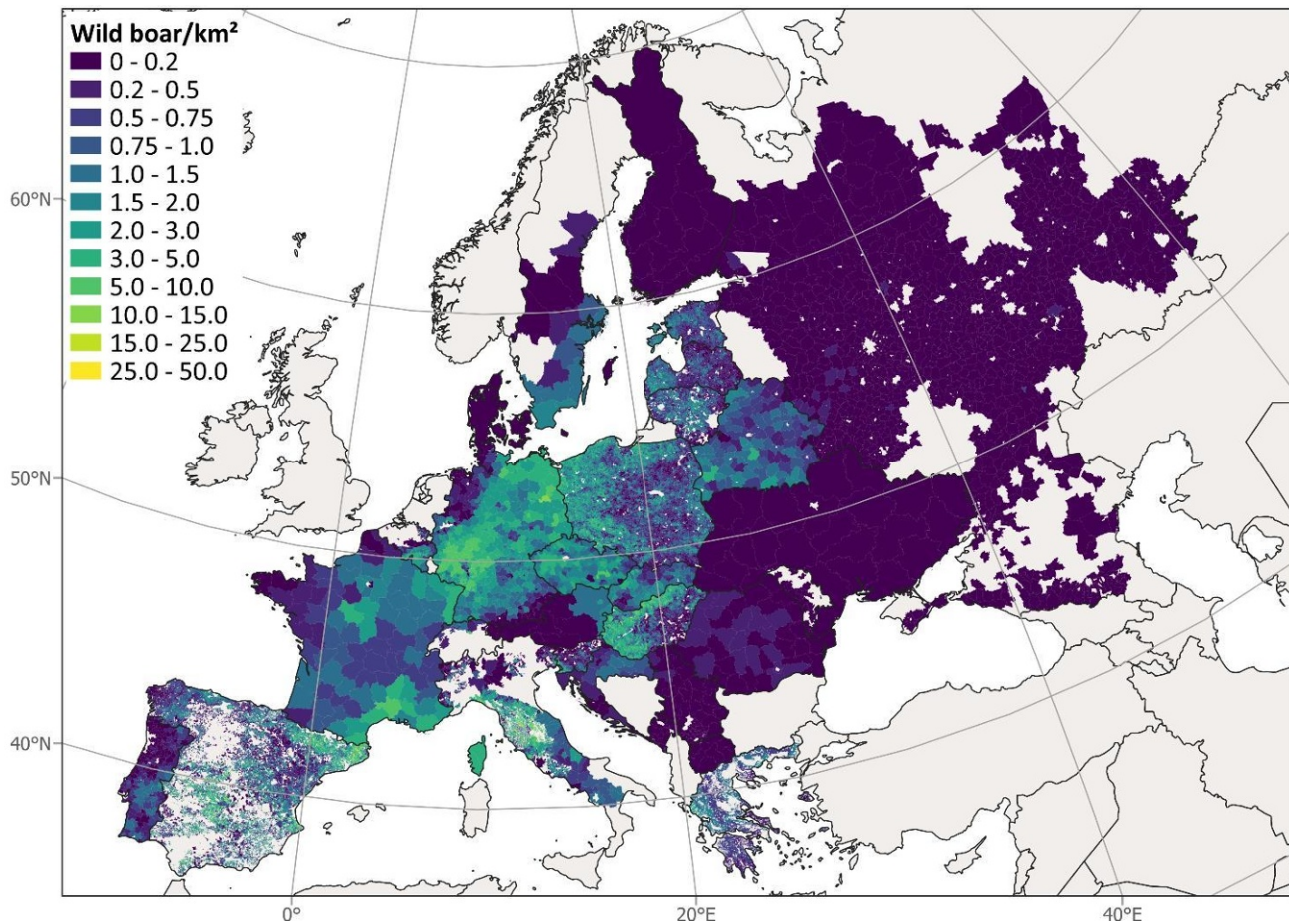


Figure 2: Hunting yield densities data used for parameterizing the models.

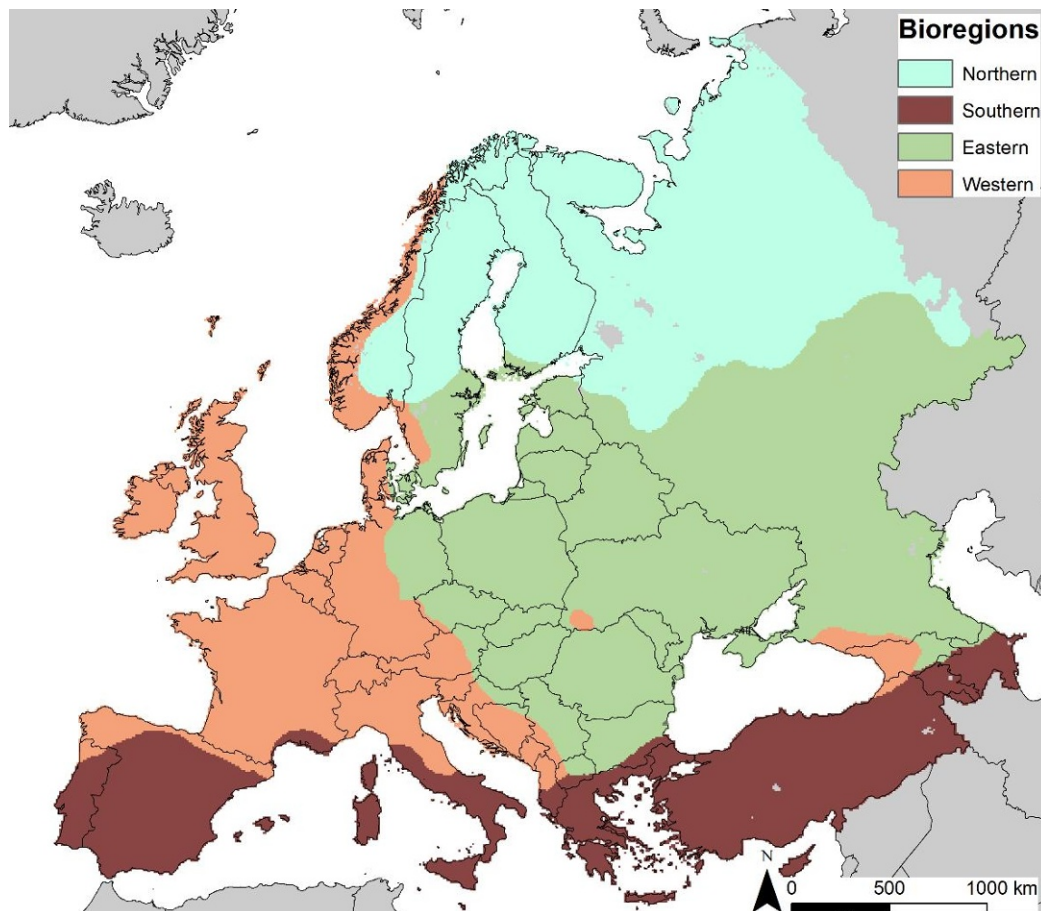


Figure 3: Map showing the bioregion classification used to subgroup the data for modelling purposes.

2.2.2. Wild ungulates occurrence data

For modelling species habitat suitability (probability of presence) we obtained occurrence records (presences) for all mammal species from the past 20 years (since 2000) exclusively from the Global Biodiversity Information Facility (GBIF); extracted on 9/12/2020 (<https://doi.org/10.15468/dl.xa9asm>). Any records without an exact sighting date, taxonomic description to the species level and coordinate accuracy equivalent to or better than that required for modelling on a 2x2 km raster grid were excluded.

When considered on a species-by-species basis, one of the biggest limitations of this type of data is the lack of information regarding survey effort which can be extremely biased both in terms of the distribution of surveyed locations and the number each is visited. Such information is critical to understand whether absence of data, in this case species presence, is evidence of true absence of a species or merely of insufficient effort to detect it. Previous studies (Phillips et al., 2009; van Strien et al., 2013; Croft et al., 2018) have suggested that records of other associated species may provide a suitable proxy to estimate survey effort. We considered records of all terrestrial mammals easily identifiable by visual observation alone (both direct and indirect, e.g., including evidence of species presence such as burrows, mounds and scat). To ensure any datasets included in our analysis complied with this idea of multi-species recording (and could therefore be considered generalised presence-effort datasets) we assessed each separately excluding any which contained observations of fewer than 10 different species or 2 taxonomic orders. Combining

all the remaining presence-only data we computed a binomial dataset for each species describing the number of successes as the visits (dates) to each study site (grid cell) where that species had been observed and the number of trials as the total number of visits (dates) to each study site where any of the species considered had been observed.

3. Methodology

3.1. Wild boar: models based on hunting yield data

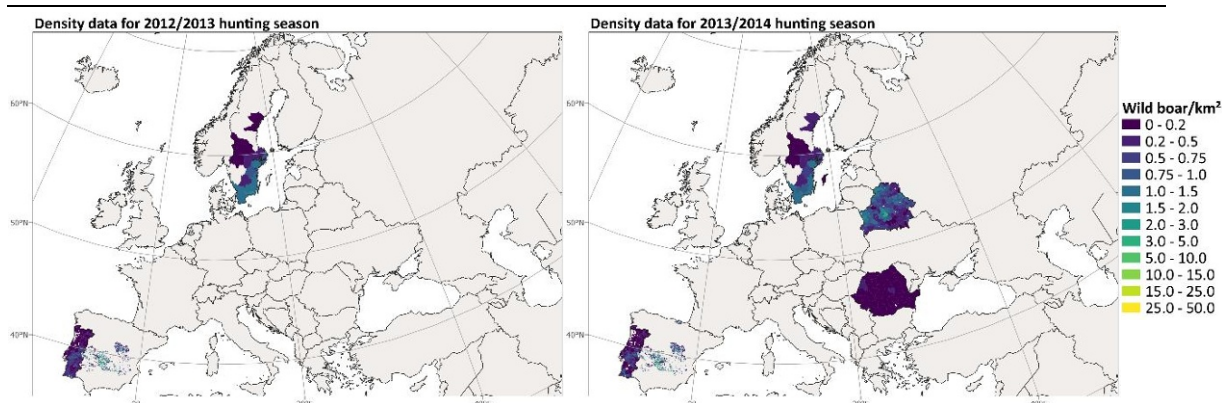
3.1.1. Temporal dataset

It is challenging to correctly manage databases when data is collected at different temporal and spatial resolutions. The recent incorporation of greater spatial resolution in the wild boar dataset allowed for the removal of previous data incorporated at lower spatial resolution. At this stage, it is not possible to analyse the temporal dataset as there were not enough data by bioregion (categorical variable) to make models by years (Table 1). Moreover, in contrast to our previous report where data covered all the European surface, at the greatest level of spatial resolution, when data collected is separated by years (hunting seasons) there are many gaps without information of wild boar in Europe (Figure 4).

To reliably analyse temporal data, is important to select the temporal data with the best spatial resolution available. This increases the amount of information needed for some variables to predict.

Table 1. Territorial unit data for Bioregion available for each year from 2012/2013 to 2018/2019 hunting season.

Hunting season	2012/ 2013	2013/ 2014	2014/ 2015	2015/ 2016	2016/ 2017	2017/ 2018	2018/ 2019
Bioregion Northern	3	3	228	245	256	256	170
Bioregion Southern	2519	2721	6306	8125	11684	6036	3244
Bioregion Eastern	11	171	968	8578	8864	8404	6112
Bioregion Western	27	58	1601	2354	2887	2707	341



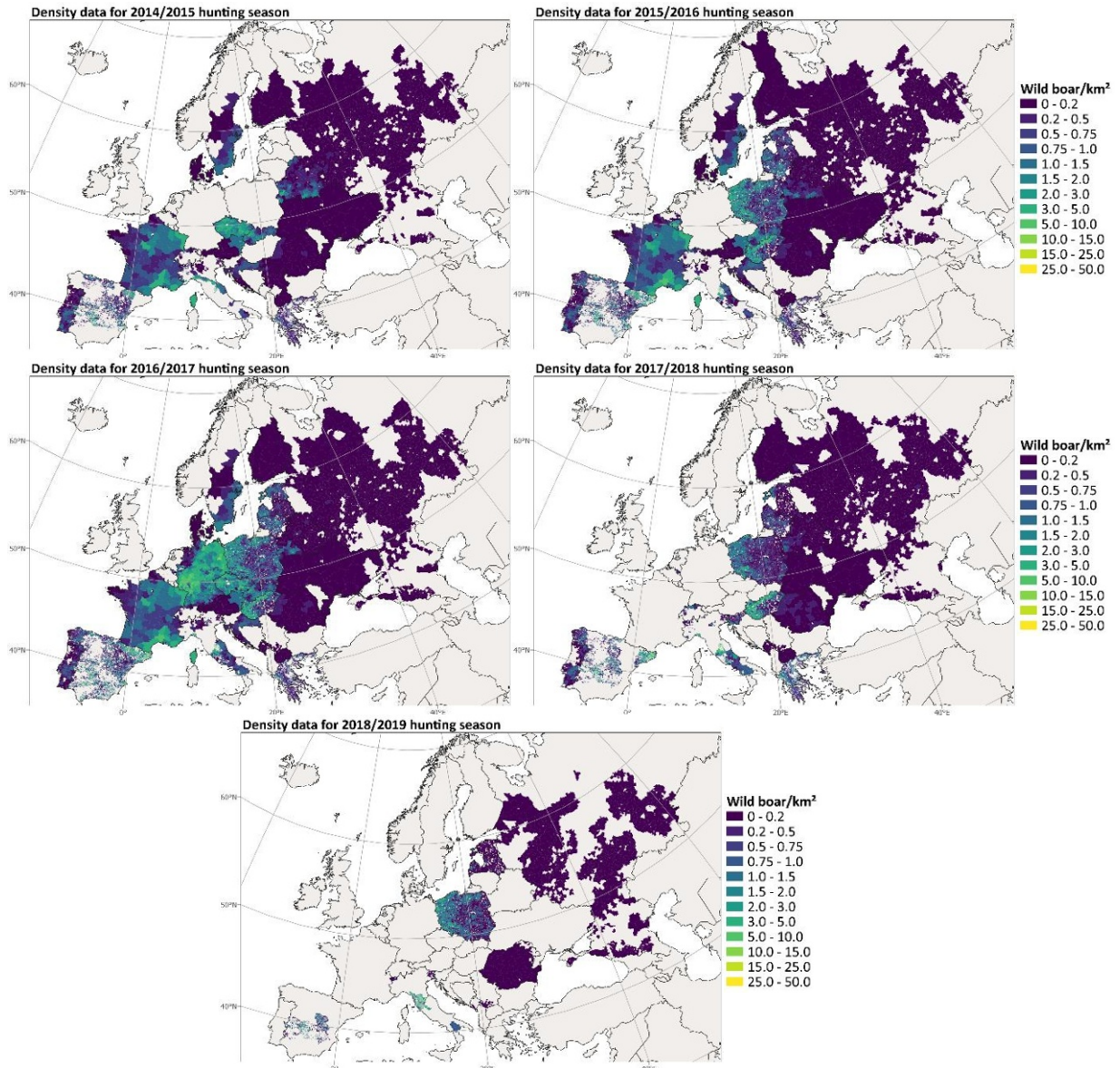


Figure 4. Available data for wild boar hunting yields for Europe from 2012/2013 to 2018/2019 hunting seasons.

3.1.2. Modelling procedure

The response variable for modelling was wild boar density (maximum number of wild boars annually hunted within 2014–2018 hunting seasons, divided by unit area in km²; hereafter HY). To allow more flexibility in predictor responses, as in previous reports, independent models were conducted for each bioregion (see also Acevedo et al., 2014; Pittiglio et al., 2018; ENETWILD consortium et al. 2020a). Therefore, we calibrated four independent models, including all ecogeographical variables as predictors. The following steps were the same as in previous reports. Ecogeographical predictors more relevant in explaining HY were determined using generalized linear models (negative binomial distribution and logarithmic link function; Cameron and Trivedi

2013). Multicollinearity among predictors was assessed using Variance Inflation Factor (VIF); predictors with VIF values above 2 were removed (Zuur et al. 2010).

All 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 individually for each bioregion as a way to assess the predictive performance and predictive resolution of the model along the abundance gradient overall for the study area (see Acevedo et al., 2014). 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 points should lie along the identity line (Pearce and Ferrier, 2001), where linearity of the relationship can be also evaluated.

Model predictions were statistically downscaled to 10x10km and 2x2km resolution, using EAA grid (<https://www.eea.europa.eu/data-and-maps/data/eea-reference-grids-2>). Before downscaling, a Multivariate Environmental Similarity Surface (MESS) analyses (see Elith et al. 2010) was carried out to assess to what extent the range of variables enclosed in the training dataset is represented at the new resolutions.

Finally, model predictions were evaluated against external data (external validation), i.e. the number of wild boar annually hunted provided from several sources at country level for Europe (see Table S1 in Pittiglio et al., 2018). So, for each country we relate the observed numbers with predicted ones using scatterplots; this allow us to assess the precision in the predicted values when models are downscaled to produce high-resolution predictive patterns of species abundance. Besides, as data belonging to this external dataset ranges from 2000-2011 and wild boar populations have been growing between 1.40-1.73 in the last decades (Massei et al., 2015), external data were transformed with the average of this growth rate (*i.e.* 1.565) from the year of the data belongs to 2015. Pearson correlation tests were made between model predictions at both resolution scales (*i.e.* 10x10km and 2x2km) and external data in order to assess for dependency.

3.1.3. Model comparisons

The consistency of the predicted pattern of wild boar abundance from HY models was assessed by comparing them with the two-step modelling approach (see previous report ENETWILD consortium et al., 2020a). For this purpose, we generated 500 random points for each bioregion in downscaled HY models at 10x10km resolution and extracted their values. Comparison was carried out by using Pearson's correlation for each independent bioregion and the relationship between them was plotted.

3.2. Models based on occurrence data

3.2.1. Bias correction

As a first step, we used a principal component analysis, rearranging the set of variables (Table 1) to minimise co-correlation, and applied a scaled average-eigenvalue test (Jolliffe 1972) to remove any redundancy. Models were fitted and tested considering survey sites within a geographic extent equivalent to a broad estimate of the "stable", or core, species range (expert-derived IUCN range maps) to exclude potential biases from absences because of dispersal-limitation or anthropogenic impact (Acevedo et al. 2012). Stationarity is an important condition of species distribution models which if not properly accounted for, particularly at large scales

where dispersal may be a significant barrier to species spread, can confound inference any about underlying environmental dependencies.

Like the condition of stationarity, autocorrelation across both space and time can confound model inference. We already account for co-correlation amongst explanatory variables and do not explicitly consider temporal changes. However, bias in the spatial arrangement of study sites (observations) can introduce significant autocorrelation which must be accounted for. This can be done explicitly within the model or removed prior to fitting. There are merits to both approaches but partially due to computational limitations and partially as our aim is to understand species presence in terms of environment rather than necessarily also describe the influence of historic dispersal we opt for the latter approach. To do this we stochastically thin our dataset ensuring sufficient spacing between sampled observations such that they can be considered spatially independent. We derive the optimal spacing between observations by evaluating the distance over which spatial autocorrelation acts within each of our explanatory variables using a Mantel correlogram to a maximum of 100 km. There are several suggestions of how then to use this information to inform spacing between observations. Here, we adopt a conservative approach taking the maximum distance across variables to ensure no residual autocorrelation. Inherent with any subsampling, this approach can destabilise the final model output, particularly where datasets are small (i.e., for species with relatively limited distributions). To account for this issue, we perform model fitting and testing on multiple subsamples and combine the results.

3.2.2. Model fitting

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

Modelling the ecological process, we considered the reduced set of explanatory variables from the PCA. For the observation process we only considered a constant to reflect that the number of trials was derived from other presence records whose detectability within any given cell was likely to be similarly affected by the environmental condition, therefore detectability was represented in relative terms compared to that of other species (i.e., difference in average size or general behaviour of species rather than environment). For all model parameters (coefficients) in both ecological and observation processes we used default uninformative Normal priors with a mean of zero and large variance of 1×10^6 providing a relatively flat distribution. Based on a MESS analysis, model projections were limited to exclude regions whose environmental conditions were deemed insufficiently represented by the training dataset so as not to produce unreliable prediction.

3.2.3. Model evaluation

Most traditional evaluation metrics rely on information about both presence and absence. Until now we have deliberately avoided a definition of absence as it has been unnecessary and can be problematic to infer from our data. However, based on estimates of observability from model fitting (probability of observation given presence) it is possible to suggest a visitation threshold above which, if present, we would expect a species to have been observed (e.g., at least one observation) and therefore the lack of presence could be taken to indicate true absence. For each training dataset we applied this definition using the fitted model to exclude any “uncertain” observation then performed a leave-one-out (LOO) cross validation. The resulting independent

predictions were then used to compute common metrics for predictive accuracy including AUC, TSS, sensitivity and specificity.

As well as evaluating predictive performance we also assess the transferability of the modelling process (variable retention/arrangement and fitting) by performing a MESS analysis using the complete dataset and all explanatory variables. The result highlights regions which may be represented by the model fit but may not have been factored into the variable selection, for example if all of our observations are in a particular climate (and therefore relatively constant) our model may exclude these variables as unimportant which may be justified but equally could be an important omission when transferring inference elsewhere, this is unknown.

4. Results and discussion

4.1. Wild boar: Model based on hunting yield data

The results from the HY models showed consistency with the wild boar spatial pattern of previous two-step models. After independent VIF analyses for each bioregion, we considered for modelling the variables listed in Table 3 and after a stepwise procedure the final models were obtained (see Table 4).

Table 3. Variables selection after VIF analyses for modelling wild boar abundance (Codes as in Table 1)

Environmental variables selected for models	
Northern Bioregion	lc_10+lc_11+lc_30+lc_40+lc_61+lc_71+lc_80+lc_110+lc_120+lc_122+lc_130+lc_160+lc_180+lc_190+lc_202+lc_210+grow+Euclidean+alt+bio_3+bio_9+bio_13+area_km2
Southern Bioregion	lc_10+lc_12+lc_20+lc_30+lc_40+lc_60+lc_70+lc_90+lc_100+lc_110+lc_120+lc_122+lc_130+lc_150+lc_153+lc_180+lc_190+lc_200+lc_201+lc_202+lc_210+grow+Euclidean+hfp+alt+snow+sun+bio_3+bio_9+bio_19+area_km2
Eastern Bioregion	lc_12+lc_20+lc_30+lc_40+lc_60+lc_61+lc_70+lc_71+lc_80+lc_90+lc_100+lc_110+lc_122+lc_130+lc_150+lc_153+lc_160+lc_180+lc_190+lc_201+lc_210+grow+Euclidean+alt+bio_8+bio_9+bio_15+area_km2
Western Bioregion	lc_10+lc_11+lc_12+lc_20+lc_30+lc_40+lc_61+lc_70+lc_71+lc_80+lc_90+lc_100+lc_110+lc_120+lc_122+lc_130+lc_150+lc_152+lc_153+lc_180+lc_190+lc_200+lc_201+lc_202+lc_210+lc_220+hfp+grow+Euclidean+bio_2+bio_8+bio_15+bio_18+area_km2

Table 4: Statistical parameters for the final models (one per bioregion) obtained to explain variation in wild boar density in Europe. Codes as in Table 1.

Model	Formula
Northern bioregion	$\text{dens} \sim \exp(-7.978e+00 + 6.914e-01 * \text{grow} + 2.831e-01 * \text{lc}_61 + 2.333e-01 * \text{bio}_3 - 1.669e-03 * \text{alt} + 6.866e-02 * \text{bio}_9 + 1.349e-02 * \text{lc}_10 - 3.212e-02 * \text{lc}_130 - 1.051e+01 * \text{lc}_122 + 1.145e-01 * \text{lc}_190)$
Southern bioregion	$\text{dens} \sim \exp(7.416e00 - 1.244e-01 * \text{bio}_3 + 1.534e-04 * \text{sun} + 1.035e-02 * \text{lc}_100 - 8.110e-02 * \text{snow} + 1.372e-02 * \text{lc}_60 - 1.988e-03 * \text{bio}_19 + 1.229e-02 * \text{lc}_70 + 9.940e-03 * \text{lc}_40 - 3.914e-02 * \text{Euclidean} - 5.805e-04 * \text{area_km2} + 7.492e-03 * \text{lc}_120 + 9.974e-03 * \text{lc}_30 + 1.231e-02 * \text{bio}_9 + 6.274e-03 * \text{lc}_20 + 3.327e-02 * \text{lc}_90 - 2.461e-02 * \text{lc}_200 - 1.937e-04 * \text{alt} - 3.493e-02 * \text{lc}_190 + 4.645e-02 * \text{lc}_122 + 5.840e-03 * \text{lc}_10 + 3.596e-03 * \text{lc}_130 + 1.866e-03 * \text{lc}_12 - 5.059e-01 * \text{lc}_202 + 2.148e-03 * \text{hfp} + 2.113e-02 * \text{lc}_180)$
Eastern bioregion	$\text{dens} \sim \exp(3.338e+00 + 2.515e-01 * \text{bio}_9 + 1.837e-02 * \text{lc}_60 + 5.793e-02 * \text{lc}_100 - 1.295e-04 * \text{area_km2} - 7.672e-03 * \text{lc}_30 - 1.086e-01 * \text{lc}_40 + 3.487e-03 * \text{lc}_70 - 1.727e-01 * \text{lc}_20 + 1.498e-02 * \text{bio}_15 + 8.682e-03 * \text{lc}_90 + 3.610e-02 * \text{bio}_8 - 1.003e+02 * \text{lc}_122 - 3.559e-02 * \text{lc}_120)$

Western bioregion

$$04*alt -1.556e-02*lc_{180} +8.840e-03*lc_{210} -3.575e+02*lc_{153} -7.349e-01*lc_{201} -2.326e+01*Eucmean +3.108e+01*lc_{71} +3.690e-03*lc_{12}$$

$$dens \sim \exp(7.887e+00 -5.804e-03*bio_{18} -4.924e-02*bio_{15} -2.178e-02*lc_{11} -1.584e-02*lc_{100} -9.799e-02*Eucmean +5.237e-02*bio_{8} -1.178e-02*lc_{130} -2.147e-02*lc_{190} -1.586e-01*lc_{201} -9.552e-05*area_{km2} -3.892e-03*lc_{70} -2.366e-01*lc_{80} -1.493e-02*lc_{20} +3.647e-02*grow -1.726e-02*lc_{12} -2.135e-02*lc_{210} -2.336e-01*lc_{220} -3.536e-01*lc_{202} -5.782e-03*lc_{90} +1.535e-02*lc_{120} -7.061e-03*lc_{30} -5.863e-02*lc_{180} +4.570e-03*lc_{10} -1.611e-01*lc_{61} -2.262e+01*lc_{152} -5.053e-01*lc_{122} -3.046e-02*bio_{2})$$

Models were then projected to the 10x10km and 2x2km grid for the entire study region. HY models' predictions suggested a similar pattern to those observed in two-step in some regions, but there are differences where highest HY predictions were reached (Figure 5).

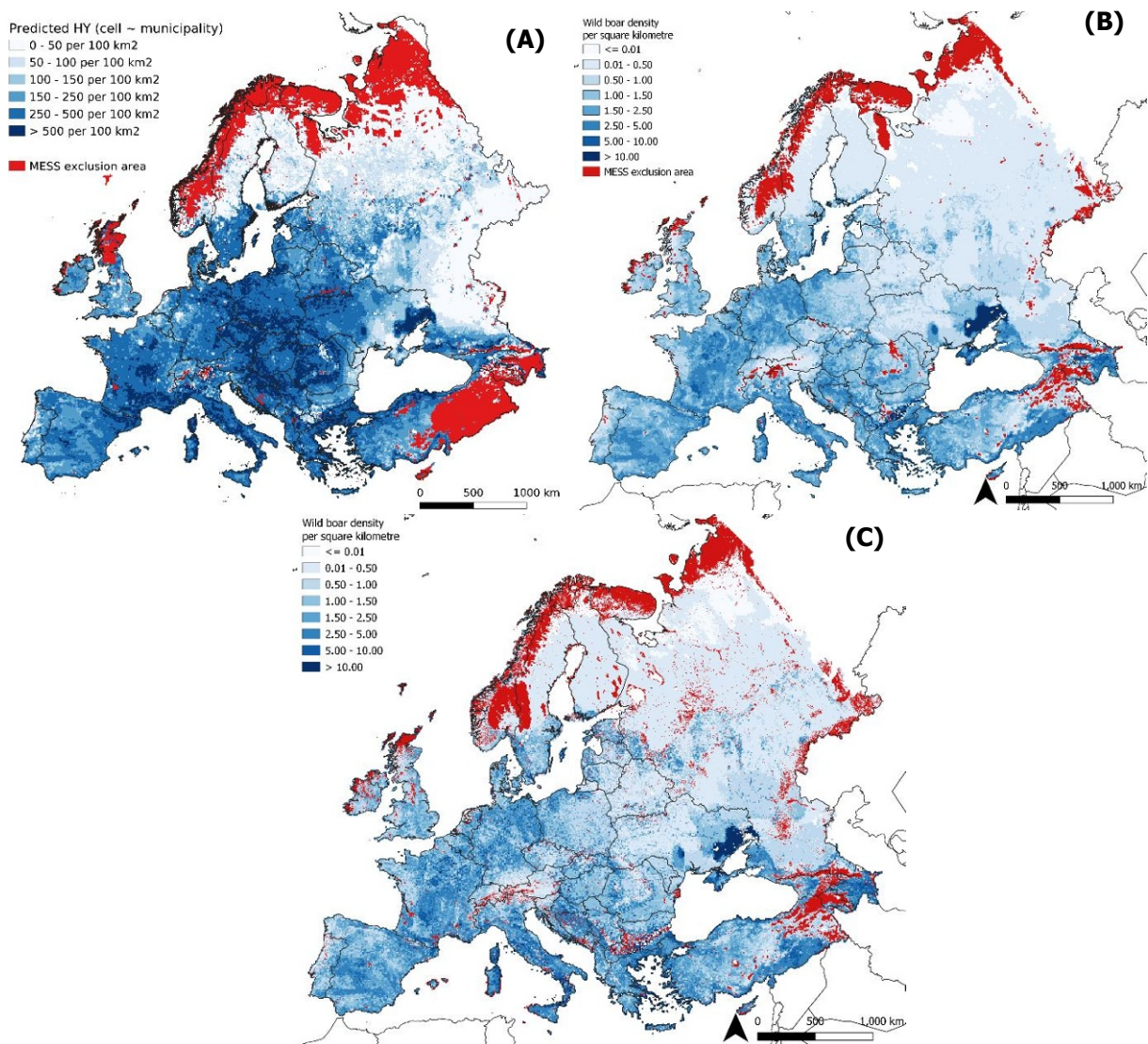


Figure 5: Hunting yield (HY) predicted at 10x10km EEA grid from the two-step modelling approach by ENETWILD consortium et al. (2020a) (A); density of wild boar hunted at 10x10km (B) and 2x2km (C) obtained in the present report. Red squares are beyond the environmental domain of the model according to MESS analyses.

The highest densities predictions were reached at Central-Western Europe, specifically in an axis delimited between France and Germany. Consecutively to this axis, densities drop off to medium values in border territories as Poland, Czech Republic, but fell to low values in Switzerland, Austria, and Slovenia. There are other axes where high densities are mixed with medium values: one between Czech Republic-Slovakia-East Austria passing through East Croatia, Bosnia and Herzegovina, Serbia, Montenegro, Bulgaria, and North-East Greece; other in central Italy, and finally, more scattered, in different mountain ranges of Iberia. Northern distribution limit was established also at 60 degrees latitude in Fennoscandia, and it is remarkable the moderate to high-density values predicted for southern Sweden.

In general, low number of wild boar densities are presented in the same regions as two-step modelling approach. These regions were characterized by alpine habitat (Alps, axial Pyrenees, and Carpathians), as well as around urban areas for East and Central Europe. Medium to low values are predicted for South bioregion, although they become locally abundant at some specific points (south of Pyrenees Mountains, East Greece / South Bulgaria) or regions (Apennines).

Model projection at greater resolution manifested slightly higher local values while lower resolution indicated a smoother pattern where the highest densities predictions are reached. The exclusion area (MESS) incremented when the model was projected at greater resolution, as differences between environmental variables are more propitious to happen when the difference between resolution of the data received for modelling and projection increases.

The calibration plot for the wild boar density approach showed a good predictive performance. Few cell values at Eastern bioregion are out of range, with values above 100 wild boar/km², showing unequivocally overprediction at local scale. Although, occasional exceptional cell values made the last quantile class very heterogeneous, which is why the ninth class of Eastern bioregion correlation plot is not shown at Figure 6, but the remaining classes fit properly between observed and predict validation data.

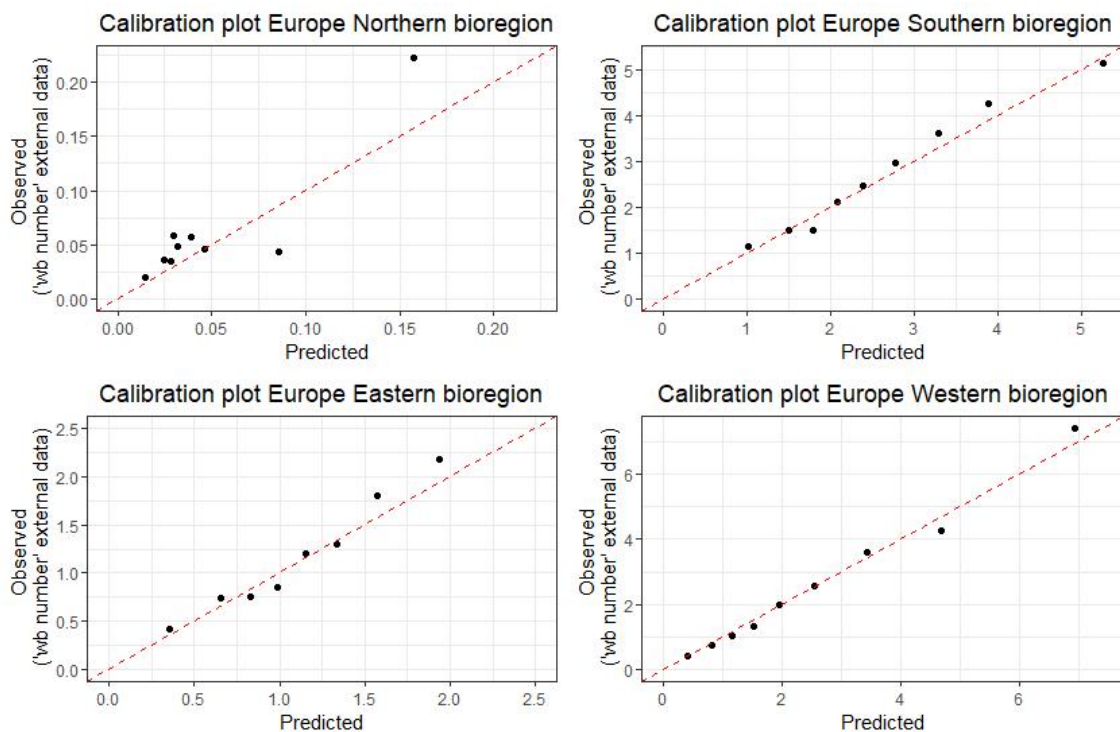


Figure 6: Calibration assessment of wild boar density modelling approach (Table 4). Plots show the relationship between the predicted wild boar densities at each bioregion and the observed wild boar densities on the validation datasets.

Model external validation of aggregated predictions at country level for European countries still showed an over-prediction of hunting bags for both model projections (10x10km and 2x2km), but there was a reduction of wild boar HY amount predicted with respect to previous models (Figure 8 and Figure 9; see ENETWILD consortium et al., 2020a), indicating that the density approach substantially reduced over-prediction.

Table 5: Pearson correlation coefficients between HY model projections at 10x10km and 2x2km and values from external datasets. Significance: ns>0.1, * <0.05, ** <0.01 and ***<0.001.

	Pittiglio et al. 2018 Not transformed	Pittiglio et al. 2018 to 2015
10x10km projection	R: 0.4978**	R: 0.4909**
2x2km projection	R: 0.5257***	R: 0.5322***

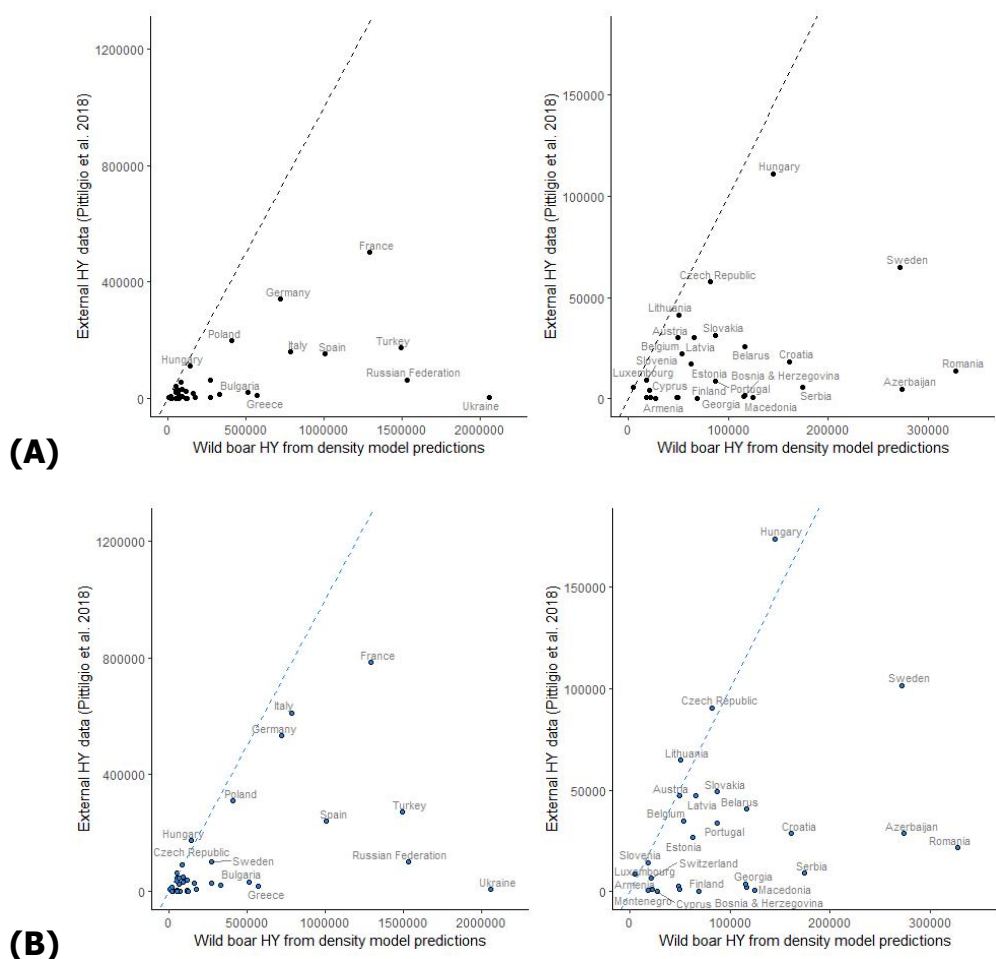


Figure 8: Relationships between aggregated model predictions at 10x10km grid cells and external validation datasets at European country level from Pittiglio et al., (2018) not transformed data (A), and data transformed to 2015 (B). Left plots show comparisons for all Europe while right plots zoomed their left below square. Dash lines show the identity ($y=x$) for each comparison.

The increase of wild boar population could explain the differences between our model predictions and external HY data, as external data back from 2000-2011 are calibrated against the maximum hunting yield reported between 2014-2018 as a response variable. When external data are increased by the wild boar population growth rate, predictions at country level for European countries show that the density approach reduces over-prediction (Figure 8B and Figure 9B), making the predicted quantities more reliable, although under-predictions have appeared also in some regions. In addition, model predictions assumed all the territory was available for hunting, including natural reserves or urban areas where hunting activity is banned. Despite this overestimation, linearity between external wild boar harvest and predictions allowed us to interpret our model as a good proxy of wild boar abundance index.

Comparisons between the previous model from ENETWILD consortium et al. (2020a) and the HY density approach showed consistence in predictions at 10x10km grid (Figure 5).

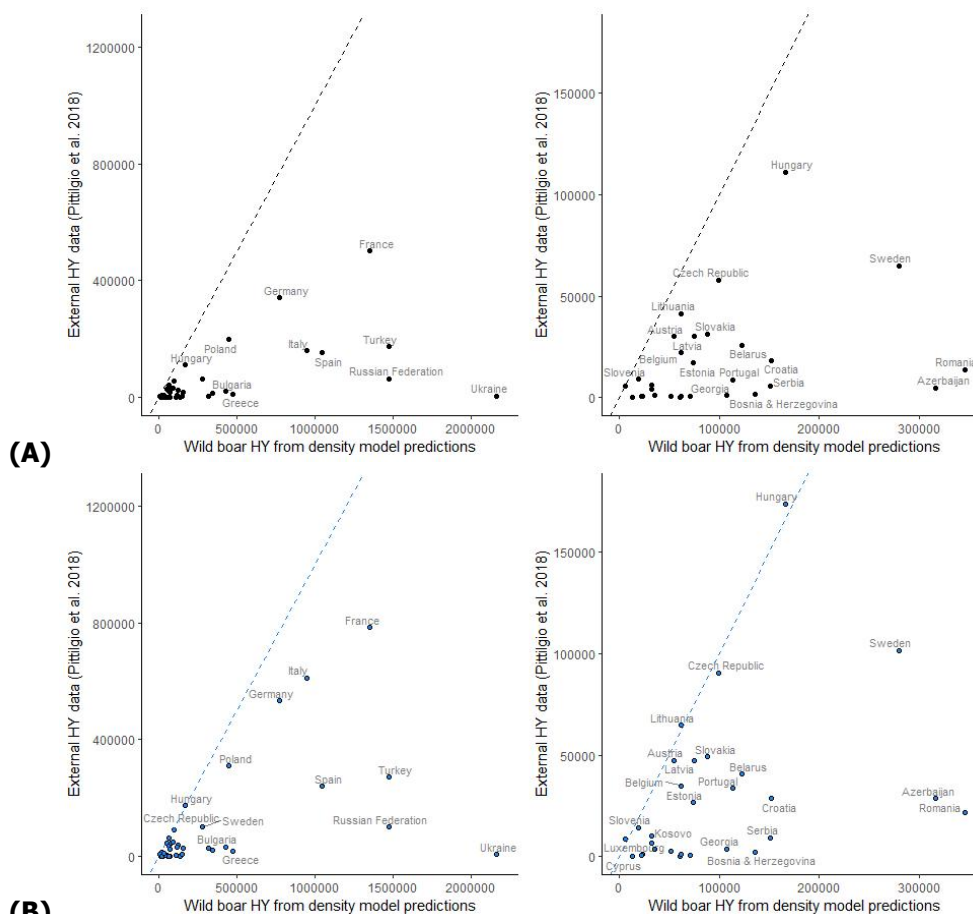


Figure 9: Relationships between aggregated model predictions at 2x2km grid cells and external validation datasets at European country level from Pittiglio et al., (2018) not transformed data (A), and data transformed to 2015 (B). Left plots show comparisons for all Europe while right plots zoomed their left below square. Dash lines show the identity ($y=x$) for each comparison.

Significant relationship between HY and two-step model predictions were observed for all bioregions (Figure 10), although the correlation strength varied among them.

The Eastern area showed the lowest Pearson' R scores ($R=0.379$; Figure 10). The low correlation value in this region was expected as it had also correlated badly with the validation dataset. The Northern area also showed a low Pearson' R score ($R=0.450$; Figure 10), although this region encompasses the northern limit of wild boar. These low correlation values could be influenced by the heterogenous resolution of biological data. Higher Pearson' R scores between the density approach and HY models were obtained for the other bioregions (Figure 10), which indicated a greater agreement in the spatial pattern predicted by density and HY models for Western Europe. General concordance between models is higher than those in previous reports (ENETWILD consortium et al., 2020a).

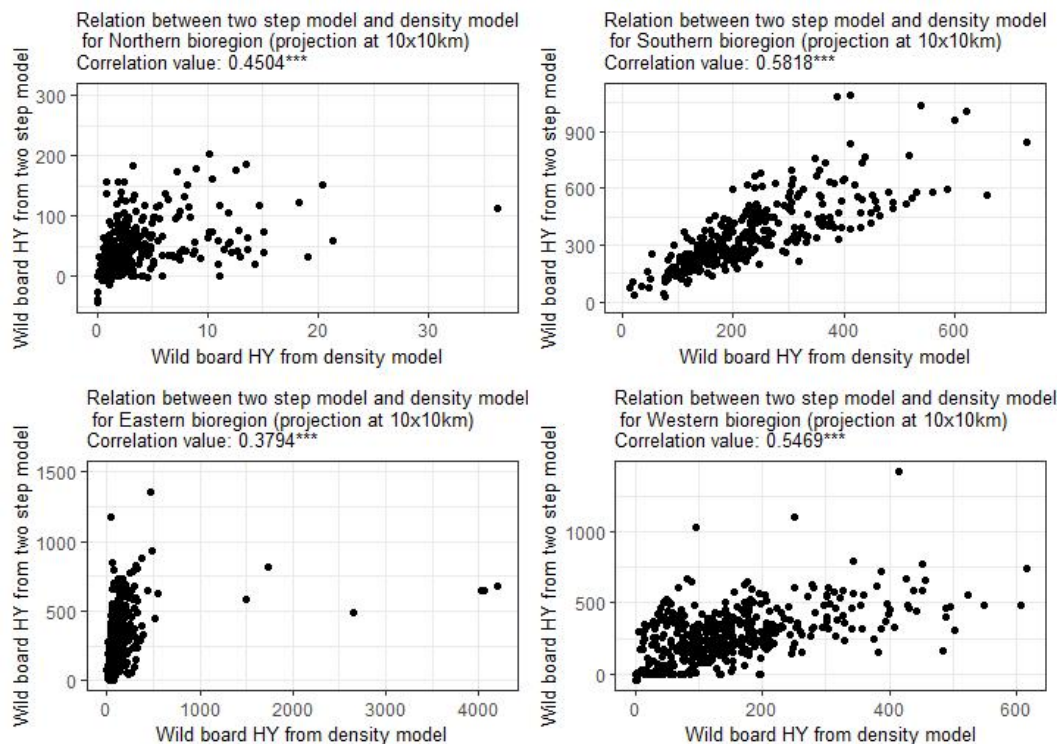


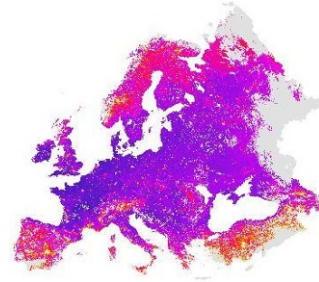
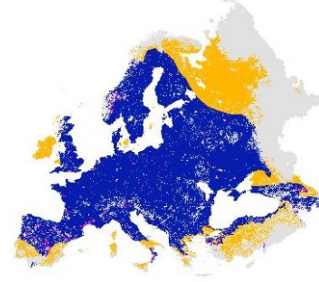
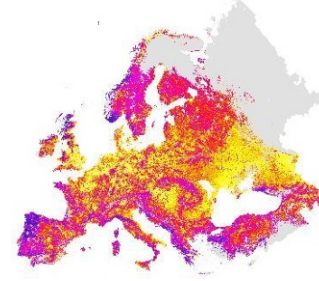
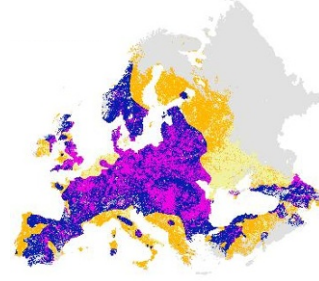
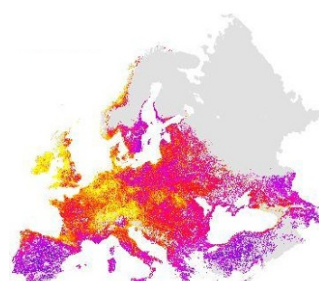
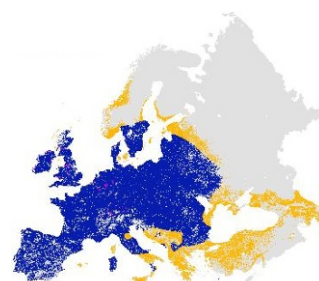
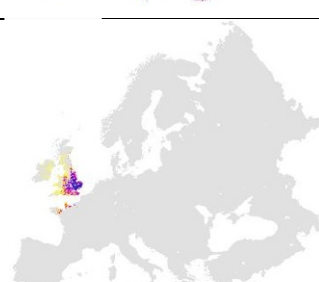



Figure 10: Relationship between wild boar HY from the two-step individual bioregion modelling approach and from density model, and Pearson's correlation coefficient for 10x10 projected resolution.

4.2. Wild ruminants and wild boar: occurrence data model

Mapped predictions from the model showing habitat suitability and corresponding occurrence for a selection of the most common "wild" (excluding feral livestock such as sheep, goats and cattle) ungulate species are provided in Figure 11. Evaluation statistics assessing model performance are shown in Table 4. Overall, these results suggest generally good model performance with AUC statistics above or close to 0.7. For those species where AUC was below 0.7 we note lower accuracy in predicting absences. The reasons for this can be difficult to disentangle as our definition of absence used for testing is itself part of the model assumptions rather than based on "true" observations and therefore may indicate inconsistencies in evaluation as much as problems with the model performance. Further work is required to investigate this issue.

	Suitability	Occurrence	Notes
<i>Alces alces</i>	<p>High: 1 Low: 0</p> <p>Habitat suitability (red to blue) Insufficient experience (grey)</p> 	<p>Absent (beyond range) (yellow) Present (within range) (dark blue)</p> <p>Present (beyond range) (orange) Insufficient experience (grey)</p> <p>Absent (within range) (purple)</p> 	<p>Distribution seems plausible. Environmental coverage of assumed "stable" range limits transferability to southern Europe. From estimate species close to realising full ecological range within Europe.</p>
<i>Capreolus capreolus</i>			<p>Predicted distribution within the UK, Western and Central Europe suggests relatively ubiquitous presence within current range with some potential to expand.</p>
<i>Cervus elaphus</i>			<p>Plausible distribution within the UK.</p>
<i>Dama dama</i>			<p>Similar to roe deer threshold for present appears a little too low. Predictions suggest species is more ubiquitous than it is particularly in the UK (where fallow are typically favour southern regions) and Western Europe as indicated by the suitability map. Observability is generally quite low so true absences are rare.</p>
<i>Muntiacus reevesi</i>			<p>Predictions would suggest species is reaching the limit of its potential distribution in the UK. Information predict suitability/potential occurrence in mainland Europe is not sufficient based on current records.</p>

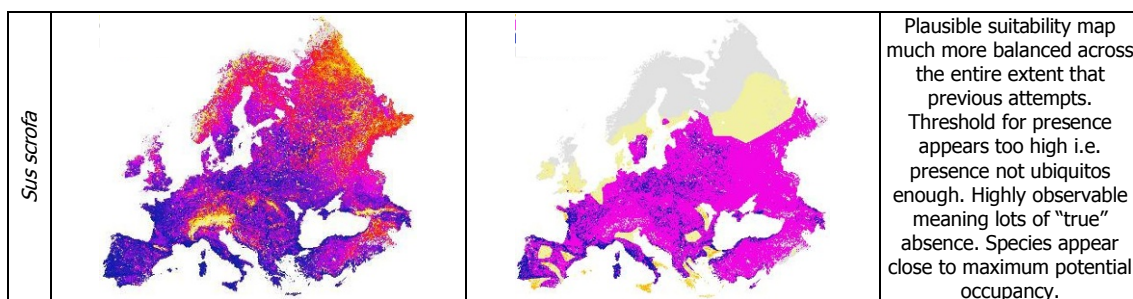


Figure 11: Predictions of wild ruminants and wild boar occurrence data models. Mapped output showing projected habitat suitability (mean across repetitions ignoring individual exclusions where model transferability may be limited) and threshold occurrence (based on mean habitat suitability and estimated threshold across all repetitions; 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.

Table 6: Mean evaluation metrics computed across repetitions for each species (AUC: Area Under the Curve; THD: Total Harmonic Distortion; SE: Sensitivity; SP: Specificity; TS: True Skill Statistic).

Species	AUC	THD	SE	SP	TSS
<i>Alces alces</i>	0.64	0.82	0.75	0.56	0.31
<i>Capreolus capreolus</i>	0.63	0.63	0.70	0.57	0.27
<i>Cervus elaphus</i>	0.72	0.35	0.75	0.71	0.46
<i>Dama dama</i>	0.71	0.49	0.66	0.84	0.50
<i>Muntiacus reevesi</i>	0.86	0.45	0.87	0.70	0.57
<i>Sus scrofa</i>	0.69	0.73	0.73	0.65	0.38

4.3. Conclusions and further steps

In relation to the modelling approach the following can be concluded:

- By modelling density of hunted wild boar rather than hunting yields, models are more reliable to predict the spatial pattern of wild ungulate species abundance at the European scale.
- Including the area of the territorial units within the response variable (by modelling wild boar density) parameterized models are more transferable among spatial scales and over-prediction when models at downscaled is mostly avoided.
- Within the constraints of the available data for habitat suitability models (probability of presence) our modelling approach appears to perform reasonably well (AUC generally close to or above 0.7, note that estimates are not inflated by spatial autocorrelation which is often the case in published studies and that absences are not certain).
 - Current modelling choices do potentially reduce the usefulness of model output specifically regarding the transferability of model projections.
 - limiting model fitting to regions where species are considered "stable" as defined by the IUCN (for many species including European rabbit for example published

ranges represent the extent of native populations which are inherently more restrictive than their naturalised "stable" range severely impacting model transferability).

Next steps for data acquisition and modelling:

- Data about wild boar abundance at the highest resolution for each country are scarce and patchy. Therefore, a complete compilation of data for different spatial resolution should be carried out for modelling temporal dimension of wild boar abundance patterns.
- Temporality in the wild boar abundance model parameterization and predictions should be explored.
- In the last report with simulations and regional data we showed that accounting for spatial autocorrelation the predictive performance of the models is improved. Feasible approaches to manage spatial autocorrelation at European scale should be assessed.
- There are several alternative approaches and improvements for habitat suitability modelling (probability of presence) which could be adopted to mitigate this issue in the general framework, applying a more data-driven approach to geographically delimit model extents.
- It will be important to balance the transparency and speed of the generic approach against more flexible (and often argued more accurate but likely slower) species specific adaptation.

References

- Acevedo, P., Jiménez-Valverde, A., Lobo, J. M., & Real, R. 2012. Delimiting the geographical background in species distribution modelling. *Journal of Biogeography*, 39(8), 1383-1390.
- Acevedo, P., F. Quirós-Fernández, J. Casal, and J. Vicente. 2014. Spatial distribution of wild boar population abundance: Basic information for spatial epidemiology and wildlife management. *Ecological Indicators* 36:594–600.
- Akaike, H. 1974. A new look at the statistical model identification. *IEEE transactions on automatic control* 19:716–723.
- Alexander, N. S., G. Massei, and W. Wint. 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(1), 23-40.
- Brus, D. J., Hengeveld, G. M., Walvoort, D. J. J., Goedhart, P. W., Heidema, A. H., Nabuurs, G. J., & Gunia, K. 2012. Statistical mapping of tree species over Europe. *European Journal of Forest Research*, 131: 145-157.
- Cameron, A. C., and P. K. Trivedi. 2013. *Regression analysis of count data*. Volume 53. Cambridge university press.
- Croft, S., Chauvenet, A. L., & Smith, G. C. 2017. A systematic approach to estimate the distribution and total abundance of British mammals. *PLoS one*, 12(6).
- Elith, J., M. Kearney, and S. Phillips. 2010. The art of modelling range-shifting species. *Methods in ecology and evolution* 1:330–342.
- ENETWILD-consortium, S. Croft, G. Smith, P. Acevedo, and J. Vicente. 2018. 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, Acevedo, P., Croft, S., Smith, G. C., Blanco-Aguiar, J. A., 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*, 16(8), 1674E.
- ENETWILD-consortium, S. Croft, G. Smith, P. Acevedo, and J. Vicente. 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, Pelayo Acevedo, Simon Croft, Graham Smith, Joaquín Vicente 2019c. ENETWILD modelling of wild boar distribution and abundance: initial model output based on hunting data and update of occurrence-based models. *FSA supporting publication* 2019:EN-1629. 25 pp. doi:10.2903/sp.efsa.2019.EN-1629.
- ENETWILD-consortium, Acevedo, P., Croft, S., Smith, G. C., Blanco-Aguiar, J. A., 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-Aguiar, J. A., Brivio, F., Croft, S., Fanelli, A., Fernández-Arias, A., Ferroglio E., Keuling, O., Levanič, T., Plis, K., Podgorski, T., Pokorny, B., Scandura, M., Smith, G. C., 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, G. C., Blanco-Aguiar, J. A., Fernandez-Lopez, J., Scandura, M., Apollonio M., Ferroglio E., Keuling, O., Sange, M. Zanet, S., Brivio, F., Podgorski, T., Petrovic, K., Soriguer R., Vicente. 2020c. Validation and inference

- of high-resolution information (downscaling) of ENETWILD abundance model for wild boar. *EFSA Supporting Publications*, 17(1), 1787E.
- Jiménez-Valverde, A. and Lobo, J.M. 2007. Threshold criteria for conversion of probability of species presence to either-or presence-absence. *Acta Oecologica*, 31(3), 361-369.
- Jolliffe, I. T. 1972. Discarding variables in a principal component analysis. I: Artificial data. *Applied statistics* 160-173.
- Latimer, A. M., Wu, S. S., Gelfand, A. E., Silander, J. A. 2006. Building statistical models to analyze species distributions. *Ecological Application*, 16(1), 33-50.
- Liu, C., White, M. and Newell, G. 2013. Selecting thresholds for the prediction of species occurrence with presence-only data. *Journal of Biogeography*, 40(4), 778-789.
- MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A., Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83(8), 2248-2255.
- Massei, G., Kindberg, J., Licoppe, A., Gačić, D., Šprem, N., Kamler, J., ... & Cellina, S. 2015. Wild boar populations up, numbers of hunters down? A review of trends and implications for Europe. *Pest management science*, 71(4), 492-500.
- Pearce, J., and S. Ferrier. 2001. The practical value of modelling relative abundance of species for regional conservation planning: a case study. *Biological Conservation* 98:33-43.
- Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., 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(1), 181-197.
- Pittiglio, C., S. Khomenko, and D. Beltran-Alcrudo. 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>.
- Royle, J. A. 2006. Site Occupancy Models with Heterogeneous Detection Probabilities. *Biometrics* 62:97-102.
- van Strien, A. J., van Swaay, C. A., Termaat, T. 2013. Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. *Journal of Applied Ecology*, 50(6), 1450-8.
- Vieilledent, G., Merow, C., Guélat, J., Latimer, A. M., Kéry, M., Gelfand, A. E., Wilson, A. M., Mortier, F., Silander Jr, J. A. 2014. hSDM: hierarchical Bayesian species distribution models. R package version 1.4. <https://CRAN.R-project.org/package=hSDM>.
- Zuur, A. F., E. N. Ieno, and C. S. Elphick. 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 0.8 or greater are considered an indication of good performance.
Background data	When not reliable information about species absence is available, a set of random points (background data) within the model extent is used to model species presence in relation to the overall environmental conditions present in the area, that are reflected in the background data.
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 is able to explain the response to the species to that range, but the model does not have information about how the species is able to respond in localities without that range
GBIF	Global Biodiversity Information Facility.
GLM	Generalised Linear Model.
Habitat suitability	the ability of a habitat to provide a species life requisites under current conditions.
Hunting bag	It refers to the number of animals hunted in a territory usually during a given hunting season.

Hunting yield	It is usually used to refer to a relative abundance index based on hunting bag data.
IUCN	International Union for Conservation of Nature.
MESS	Multivariate Environmental Similarity Surface.
Model extent	This term refers to the geographical area on which the model is to be fitted.
NUTS3	Nomenclature of Territorial Units for Statistics Level 3.
Population density (d)	It is a measurement of population size per area unit, i.e., population size divided by total land area. The absolute density usually is expressed in heads per 100 ha. Multiplying the population density by the studied surface, we obtain the population size. It can be calculated by different methods (either direct or indirect, summarized in Table 1).
Population size or absolute abundance (N)	It is the size of the population. It can be a known or estimated number, expressed in number of individuals. When related to area unit it gives the population density.
Predictive accuracy	Quantitative metric describing the accuracy of model predictions. Computed by comparing model predictions against independent data often obtained through a process of cross-validation.
Presence-absence	Dataset contain independent locations with binary classification describing whether a species is present or explicitly absent.
Presence-background	Independent datasets describing environmental conditions at locations where a species has been observed and those of a random sample from the available landscape.
Presence-only	Dataset containing independent events (date, location, recorder) describing species sightings; positive occurrences.
Relative abundance	Index describing the difference in populations across locations. Typically expressed using a discrete classification scale. When expressed as a continuous scale relative abundance can be transformed in absolute abundance using a population count at a single location.
Relative score	Index describing the difference in suitability, i.e. likelihood of species presence, across locations.
RSF	Resource Selection Function.
Suitability	Measure of how suitable a location is for a particular species; analogous to the likelihood that a species is present.
Training dataset	Split the dataset is a common modelling practice aimed to use a proportion of data to fit the model (training dataset) and the rest of data to assess the model performance on independent (i.e. not use in model fitting) data (evaluation dataset).
TSS	True Skill Statistic is a measure of model accuracy which considers omission and commission errors, and success as a result 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.

WBDM

ENETWILD Wild Boar Data Collection Model.