ECOSPHERE
AN ESA OPEN ACCESS JOURNAL

Importance of data selection and filtering in species distribution models: A case study on the Cantabrian brown bear

Alejandra Zarzo-Arias[1,2,3] 🄳 | Vincenzo Penteriani[4,5] 🄳 | Lukáš Gábor[3,6,7] | Petra Šímová[3] | Florencia Grattarola[3] | Vítězslav Moudrý[3]

[1]Department of Biogeography and Global Change, Museo Nacional de Ciencias Naturales (MNCN-CSIC), Madrid, Spain

[2]Universidad de Oviedo, Oviedo, Spain

[3]Department of Spatial Sciences, Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Prague, Czech Republic

[4]Department of Evolutionary Ecology, Museo Nacional de Ciencias Naturales (MNCN-CSIC), Madrid, Spain

[5]Biodiversity Research Institute (IMIB, CSIC-UO-PA), Mieres, Spain

[6]Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut, USA

[7]Center for Biodiversity and Global Change, Yale University, New Haven, Connecticut, USA

Correspondence
Vincenzo Penteriani
Email: v.penteriani@csic.es

Funding information
European Union-NextGenerationEU, Ministerio de Universidades y Plan de Recuperación, Transformación y Resiliencia, Universidad de Oviedo, Grant/Award Number: Margarita Salas; Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Grant/Award Number: 2022B0012; Spanish Ministry of Science and Innovation, the Agencia Estatal de Investigación (AEI), and the Fondo Europeo de Desarrollo Regional (FEDER, EU), Grant/Award Number: PID2020-114181GB-I00

Abstract

Species distribution models (SDMs) are powerful tools in ecology and conservation. Choosing the right environmental drivers and filtering species' occurrences taking their biases into account are key factors to consider before modeling. In this case study, we address five common problems arising during the selection of input data for presence-only SDMs on an example of a generalist species: the endangered Cantabrian brown bear. First, we focus on the selection of environmental variables that may drive its distribution, testing if climatic variables should be considered at a 1-km analysis grain. Second, we investigate how filtering the species' data in view of (1) their collection procedures, (2) different time frames, (3) dispersal areas, and (4) subpopulations affects the performance and outputs of the models at three different spatial analysis grains (500 m, 1 km, and 5 km). Our results show that models with different input data yielded only minor differences in performance and behaved properly in terms of model validation, although coarsening the analysis grain deteriorated model performance. Still, the contribution of individual variables and the habitat suitability predictions differed among models. We show that a combination of limited data availability and poor selection of environmental variables can lead to inaccurate predictions. Specifically for the brown bear, we conclude that climatic variables should not be considered for exploring habitat suitability and that the best input data for modeling habitat suitability in the study area originate from (1) observations and traces from the (2) most recent period (2006–2019) in which the population is expanding, (3) not considering cells of dispersing bear occurrences and (4) modeling subpopulations independently (as they show distinct habitat preferences). In conclusion, SDMs can serve as a useful tool for generalist species including all available data; still, expert evaluation from the perspective of data suitability for the purpose of modeling and possible biases is recommended. This is especially important when the results are intended for management and

conservation purposes at the local level, and for species that respond to the environment at coarse analysis grains.

INTRODUCTION

Exploring how species respond to different environmental characteristics is key to creating successful management and recovery plans (Jetz et al., 2012). Species distribution models (SDMs) are currently popular tools for this purpose, especially considering the increase in the occurrence data availability (Guisan et al., 2013; Rodríguez et al., 2007). As an example of this trend in animal and plant modeling, we can name the increasingly used MaxEnt. This method based on the maximum-entropy approach (Merow et al., 2013; Phillips & Dudík, 2008) has become popular over the last decade because it can work with presence-only records of species' occurrences that are increasingly available from global databases (Jayathilake & Costello, 2020; Kuralt & Kos, 2018; Zhang et al., 2021). There are, however, key factors that need to be considered by researchers before modeling, such as choosing the right environmental drivers (Gardner et al., 2019) or filtering available species' occurrences and coping with their biases, for example, sampling bias or positional error (Fernandes et al., 2018; Gábor, Moudrý, Lecours, et al., 2020; Kramer-Schadt et al., 2013; Moudrý & Šímová, 2012). If not, SDMs may provide misleading results and conclusions, which can be particularly problematic when applied to population management and conservation.

SDMs relate species' occurrences to the environmental characteristics of their habitat and assess potential landscape–species relationships and geographical range suitability (Elith et al., 2011). Thus, the correct selection of environmental characteristics that may play a role in the distribution patterns of target species is one of the crucial modeling steps. Not much attention has been given to the selection of variables and there is little consistency among studies regarding the used variables (Austin & Van Niel, 2011; Fourcade et al., 2018; Gábor et al., 2022). Some studies even rely only on climate-derived variables; this is especially true for studies predicting species shifts due to climate change, although the climate may not be enough to predict major shifts in species' distribution (Araújo & Guisan, 2006). Especially in vertebrates, such as large mammals, climatic variables might influence some species only

indirectly by affecting their food and shelter as their habitat selection would be more likely driven by landscape features and structure and/or human impact (Berteaux & Stenseth, 2006; Greenberg et al., 2014). Thus, many studies fail at choosing important variables shown to be physiologically relevant to the species (Gardner et al., 2019).

Furthermore, species occurrences mostly come from non-systematic observations from museum collections, administrations, or citizen science, usually accessible from global public databases, such as Global Biodiversity Information Facility (GBIF, www.gbif.org) (Moudrý & Devillers, 2020; Newbold, 2010). Information from these sources can be affected by several (generally more than one) errors related to the observer, and the tools and/or collection methods employed (Marcer et al., 2022; Moudrý & Devillers, 2020). For example, data might suffer from positional uncertainty/inaccuracy, sampling effort bias (i.e., sampling toward places where it is easy to locate individuals or that are highly visited, but not from areas with reduced visibility or more remote and inaccessible), or different ecological meaning (i.e., observations recorded where a specific behavior takes place, such as locations of damages to human property or of water sources) (Anderson, 2012; Fourcade et al., 2014; Guillera-Arroita, 2017; Kramer-Schadt et al., 2013). Thus, the proper selection of species' input records is a necessary step toward construction of accurate habitat models (Beck et al., 2014; Boria et al., 2014). Common approaches to reduce the impact of bias in occurrence data include, for example, spatial or environmental filtering (Gábor, Moudrý, Barták, & Lecours, 2020; Inman et al., 2021; Varela et al., 2014; Veloz, 2009), subsampling (Beck et al., 2014), or the use of systematic sampling or a bias file to account for sampling bias and to weight occurrences accordingly (El-Gabbas & Dormann, 2018; Fourcade et al., 2014; Tessarolo et al., 2014). Still, performing a careful revision of the available data to evaluate and understand various errors and biases in view of the data sources or collection methods before modeling is uncommon. Ignoring the intrinsic inaccuracies and ecological differences of the data can, however, produce misleading results and conclusions (Guillera-Arroita, 2017; Isaac & Pocock, 2015; Leitão et al., 2011).

Still, the sources of possible bias are broader than just "how and by whom the data are collected." Long-term monitoring information available for some species has the potential to provide quality data for studying species distributions (Brotons et al., 2007). However, different periods might reflect different population statuses associated with density-dependent use of space and, consequently, different habitat requirements (Seoane & Carrascal, 2008). Besides, data collected more recently are likely to be more accurate in their geolocation due to the wide availability and ongoing improvement of the GPS technology (Tomaštík & Varga, 2021). Population status should also be considered when filtering species data as, for example, dispersing individuals in expanding populations can appear in unfavorable areas while searching for new landscapes or moving between suitable habitat patches; their habitat selection may considerably differ with their breeding status as well (Moudrý et al., 2017; Rayment et al., 2015; Stamps, 2001). Additionally, different subpopulations or clusters might differ in the (1) growth/reproductive rate, (2) habitat or resources availability, (3) behavior, (4) impact of human infrastructure and density, (5) size/number of individuals (which might affect density-dependent patterns), or (6) number of occurrences or sampling effort. All of these can influence the models' behavior. Models may, therefore, be improved by splitting the species into subunits with different ecological characteristics or with different sampling efforts (Fourcade et al., 2014; Stockwell & Townsend Peterson, 2002).

Faced with the many problems of ecological modeling, a question arises: Is it better to use as much data as possible, disregarding their error or ecological meaning, or should we try to screen the data and minimize the bias? This is a key question that should be acknowledged before modeling species distributions and considered when building models, interpreting the results, and drawing conclusions (Bloom et al., 2018; Guillera-Arroita et al., 2015; Kramer-Schadt et al., 2013). In this study, we intended to test the importance of selecting and filtering proper input data (i.e., variables and occurrence data) to build SDMs in a case study for a generalist species, as well as to make these decisions based on expert knowledge of the analyzed species and populations. For this, we applied MaxEnt modeling to the distribution data on the endangered Cantabrian brown bear (*Ursus arctos*) population located in northwestern Spain from 1985 to 2019. The small number of bears in this population (Pérez et al., 2014), their complete isolation, and the limited connectivity between subpopulations make this population an interesting model in terms of conservation and management. We hypothesized that (1) climatic variables would not correctly define bear habitat selection as this species is dependent rather on landscape features and inhabits a wide range of climatic regions all around the world (Krofel et al., 2021; Penteriani et al., 2021) and (2) that thorough filtering of species' occurrences will produce better models relying on both model validation and expert opinion.

METHODS

Study area and species

In our study, we focus on the brown bear population inhabiting the Cantabrian Mountains, located in the northwest of the Iberian Peninsula (Figure 1), a typical human-modified landscape of southern Europe (Penteriani et al., 2021). This population is divided into the western subpopulation, with 203 bears (95% CI = 168–260) in 2014 (Pérez et al., 2014) occupying an area of more than 7000 km^2, and the eastern subpopulation, with 19 bears (95% CI = 12–40) in 2014 (Pérez et al., 2014), occupying around 4000 km^2 (Lamamy et al., 2019; Penteriani et al., 2021). The long-term monitoring of the Cantabrian brown bear population, which started in the 1980s, is essentially based on yearly direct sightings, indirect signs (traces) of presence (footprints, hairs, and scats), images from camera traps randomly located in the bear distribution range (Penteriani et al., 2019; Zarzo-Arias et al., 2019), and records of damage caused to livestock, beehives, crops, human activities, and infrastructures (Zarzo-Arias et al., 2020). Data have been gathered by the personnel of the Principado de Asturias and Junta de Castilla y León, primarily the Patrulla Oso, as well as all the other guards of both regional governments, by Fondo para la Protección de los Animales Salvajes (FAPAS), Fundación Oso de Asturias (FOA), Fundación Oso Pardo (FOP), and by the Cantabrian Brown Bear Research Group (http://www.cantabrianbrownbear.org/es/).

Data were divided into groups considering first the type of data: (1) direct and indirect (camera traps) bear observations; (2) traces (footprints, hairs, and scats); and (3) damages, as each one has different intrinsic ecological and sampling considerations. Then, we divided the data into (1) the whole time frame (1985–2019), (2) the older (1985–2005), and (3) the recent period (2006–2020). This temporal division corresponds to the main trend change observed in this bear population, which shifted from slow recovery to population increase and colonization of new areas that occurred around 2005 (Mateo-Sánchez et al., 2015; Penteriani et al., 2021; Zarzo-Arias et al., 2019). We also grouped the data according to the western and eastern subpopulations, reflecting the
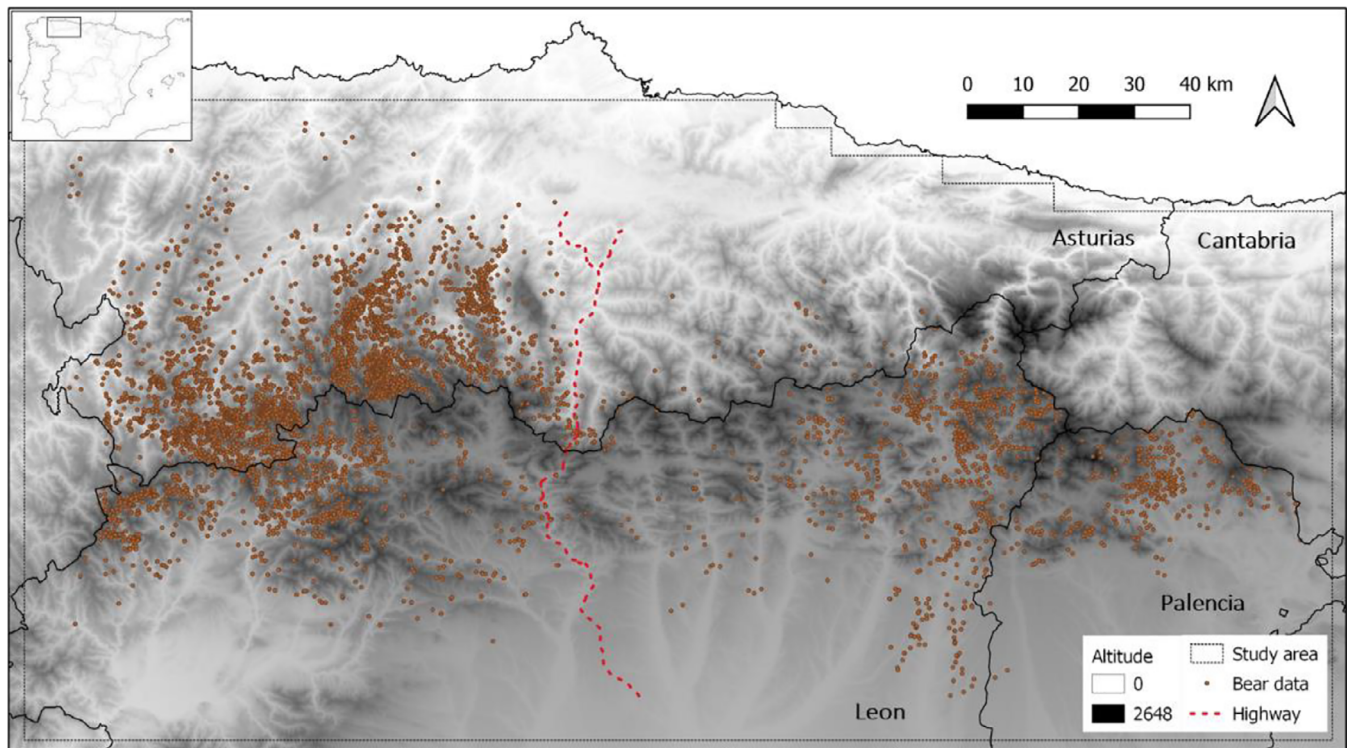
FIGURE 1 Area of study (the Cantabrian Mountains) and brown bear observations, traces, and damages (brown points) used in this study. The dotted red line indicates the highway dividing the two bear subpopulations.

highway dividing the area and acting as the main dispersal barrier (Figure 1).

Environmental data

We considered ecologically relevant environmental variables related to the terrain (5), climate (19), habitat (7), and human activity (5). Terrain, habitat, and human activity data are available for download from the Spanish National Center of Geographic Information (https://centrodedescargas.cnig.es/). The habitat variables were derived from the national Corine Land Cover database for the year 2018. For each analysis grain, we calculated the area of the following land uses to get the percentage per cell: urban areas, forests, crops, shrublands, pastures, rocks, and conifers. In addition, we calculated the distance to highways, roads, trails, and rivers. The terrain variables (Lecours et al., 2017) were derived from a high-accuracy digital terrain model at a 25-m resolution (Moudrý et al., 2018). The terrain attributes were derived using Horn's algorithm at that resolution and subsequently aggregated to the analysis grain using the mean value from each cell (Moudrý et al., 2019). The human population density obtained from Gallego (2010) was also included in the analysis. Nineteen standard bioclimatic variables were downloaded from the

CHELSA climatic portal (https://chelsa-climate.org/). We extracted the information from all variables in grid cells of 5 km, 1 km, and 500 m (i.e., analysis grain) in our study area using QGIS version 3.8 (Quantum GIS Development Team, 2015). We examined pairwise correlations of variables as was done in many other studies (Brun et al., 2020; Lecours et al., 2020; Parra et al., 2004); from those showing high mutual correlations (Pearson's coefficient >0.7), we always kept the one most relevant to the species ecology. The correlated variables included mainly climatic variables (detailed description of variables can be found in Appendix S1); the four climatic variables most relevant to the species ecology were kept (isothermality, temperature seasonality, annual precipitation, and precipitation seasonality). Subsequent variance inflation factor (VIF) analysis did not find any potential multicollinearity at any analysis grain (VIF < 5; Zuur et al., 2010). Finally, we removed variables with ≤0.1% mean contribution to a preliminary set of models at a 1-km grain. A total of 16 variables were retained for modeling (see Appendix S1).

Evaluated input data characteristics

It has been hypothesized that organisms respond to their environment more strongly at some grains than at others.

In general, it is assumed that climate defines the distribution of species at broad spatial grains. At successively finer grains and at regional extents, the importance of topography variables in controlling species occurrence increases; at even finer grains, the presence of individual land cover categories can play a role (Field et al., 2009; Pearson & Dawson, 2003). As a consequence, the choice of the analysis grain can strongly affect the ability to detect and measure species' response to the environment (Huston, 2002). As it has been shown that different variables may affect bears' distribution at different spatial grains (Mateo-Sánchez et al., 2014), we tested three different spatial analysis grains (square grids of 5 km, 1 km, and 500 m) to account for this.

To choose the optimal way to build the Cantabrian bear habitat suitability models, we first built a set of 1-km models with and without climatic variables as climate variables, together with land cover ones, which were reported to be the most important predictors of this population at finer analysis grains (see Zarzo-Arias et al., 2019), while topography does at coarser resolutions. We used the model performance and expert knowledge of the species and its distribution within the study area to evaluate whether climatic variables were driving the modeled species distribution. Then, we compared models with input data (1) of different types (observations, traces, damages, and their combinations); (2) from different time frames (1985–2005 and 2006–2019); (3) considering grids with just a single occurrence (i.e., cells used only for dispersion that do not show habitat preferences but the casual movement around the landscape looking for new suitable places) as non-presences and as presences; and (4) from the whole population and from each subpopulation independently. Subsequently, we selected the best combination of input data. Finally, to evaluate the impact of data filtering on the results, we compared this model to the one built using non-filtered data (i.e., the model containing all types of data for the entire time frame [1985–2019] using all grids with at least one occurrence for the whole population).

Model fitting and selection

For building the models, we used the software MaxEnt (Phillips et al., 2004, 2006) version 3.3.3k called from the R environment version 3.5.1 (R Core Team, 2020) using the packages dismo version 1.3.3 (Hijmans et al., 2017) and ENMeval version 0.3.1 (Muscarella et al., 2014) following the methodology applied by Zarzo-Arias et al. (2019) at three different spatial analysis grains (square grids of 5 km, 1 km, and 500 m). Input data consisted of the centroid coordinates of the cells where at least one

bear occurrence falls in. It is recommended to evaluate the best combination of MaxEnt parameter settings to select the most appropriate models (Morales et al., 2017). Therefore, to identify an optimal model structure, we evaluated candidate models with all types of feature combinations representing different types of parameterizations (linear, product, quadratic, hinge, threshold, and categorical), each run over a set of regularization multipliers ranging from 0 to 19 to control for overfitting by penalizing variables that add little to the model (see Elith et al., 2011, for more details). We used 500 iterations (maximum number of steps to reach convergence), and a convergence threshold (probability of predicting no occurrence where there is one) of 10^{-5}, using the method checkerboard1 for data partitioning. This method aggregates the original environmental input grids to partition occurrence localities into two bins for training and testing (Muscarella et al., 2014). The MaxEnt default number of background points (10,000) has been criticized as insufficient to capture the total range of environmental conditions of the targeted landscape (Renner & Warton, 2013). For this reason, we used values from all grids in the study area (specifically 1019 background cells for the 5-km^2 resolution, 25,475 for 1 km^2, and 101,900 for 500 m^2) to build all the models (as in, e.g., Ericksson & Dalerum, 2018; Swanepoel et al., 2013).

We identified the most robust and parsimonious combination of feature types and regularization multipliers using Akaike information criterion corrected for small sample sizes (AIC_c) following Warren and Seifert (2011). We considered models within two AIC_c units to have equivalent empirical support (Burnham & Anderson, 2002) and chose the simplest model (lowest number of parameters, and if equivalent, lowest number of feature types) as the best model. The model structure and the total number of grids considered for each group of bear data in each analysis grain are available in Appendix S2.

Model evaluation

It is recommended to assess model performance using multiple metrics. Therefore, we used three validation metrics in this study (Grimmett et al., 2020). We assessed the model discrimination ability (i.e., how well the model differentiates between the presence and non-presence sites) using the area under the receiver-operator curve (AUC; Fielding & Bell, 1997). AUC ranges from 0.5 for models with no discrimination ability to 1 for models with perfect discrimination. In addition, we adopted also the mean absolute error (MAE) approach recently recommended by Konowalik and Nosol (2021) to be

combined with AUC for selecting the best performing models, as the latter might be affected by large numbers of background cells (Lobo et al., 2008). The MAE of a model is the mean of the absolute values of the individual prediction errors (i.e., the difference between the true value and the predicted value) over all instances in the evaluation set. Finally, we calculated the Boyce index, which is recommended for presence-only models and shows the accuracy of the predicted probability of presence (Hirzel et al., 2006). To calculate the index, the predicted suitability is partitioned into suitability classes and, subsequently, the ratio between the predicted and expected frequency of evaluation points is plotted against the corresponding suitability class. If the model correctly distinguishes between species-suitable and unsuitable areas, the ratio between predicted and expected (random) frequencies of presences increases with the suitability (i.e., the low-suitability classes should contain fewer evaluation presences, while high-suitability classes should contain more evaluation presences than expected by chance). This increase is measured using the Spearman rank correlation coefficient, and hence the Boyce index ranges from −1 to 1. Negative values indicate an incorrect model that predicts low suitability in areas with more frequent presences, while positive values indicate a model whose suitability predictions match the distribution of presences in the evaluation dataset. We adopted the continuous version of Boyce index, which overcomes the sensitivity of the index to the number of suitability classes (Hirzel et al., 2006; Jiménez & Soberón, 2020).

After evaluating the models' predictive performance, we compared them according to the relative contribution of each environmental variable, which was calculated by a jackknife procedure and a heuristic method provided by MaxEnt. Additionally, with the raster output of the predictions, for which we used the complementary log–log format, we performed a niche comparison analysis based on the Schoener's I statistic (Warren et al., 2008), which ranges from 0 (no overlap) to 1 (perfect overlap). Finally, we compared the predicted suitability at presence locations and map outputs graphically.

RESULTS

Climatic variables

Models including climatic variables showed higher AUC but worse performance according to MAE (Figure 2a), and slightly higher Boyce index values than models without climatic variables (Appendix S2: Table S1). Climatic variables a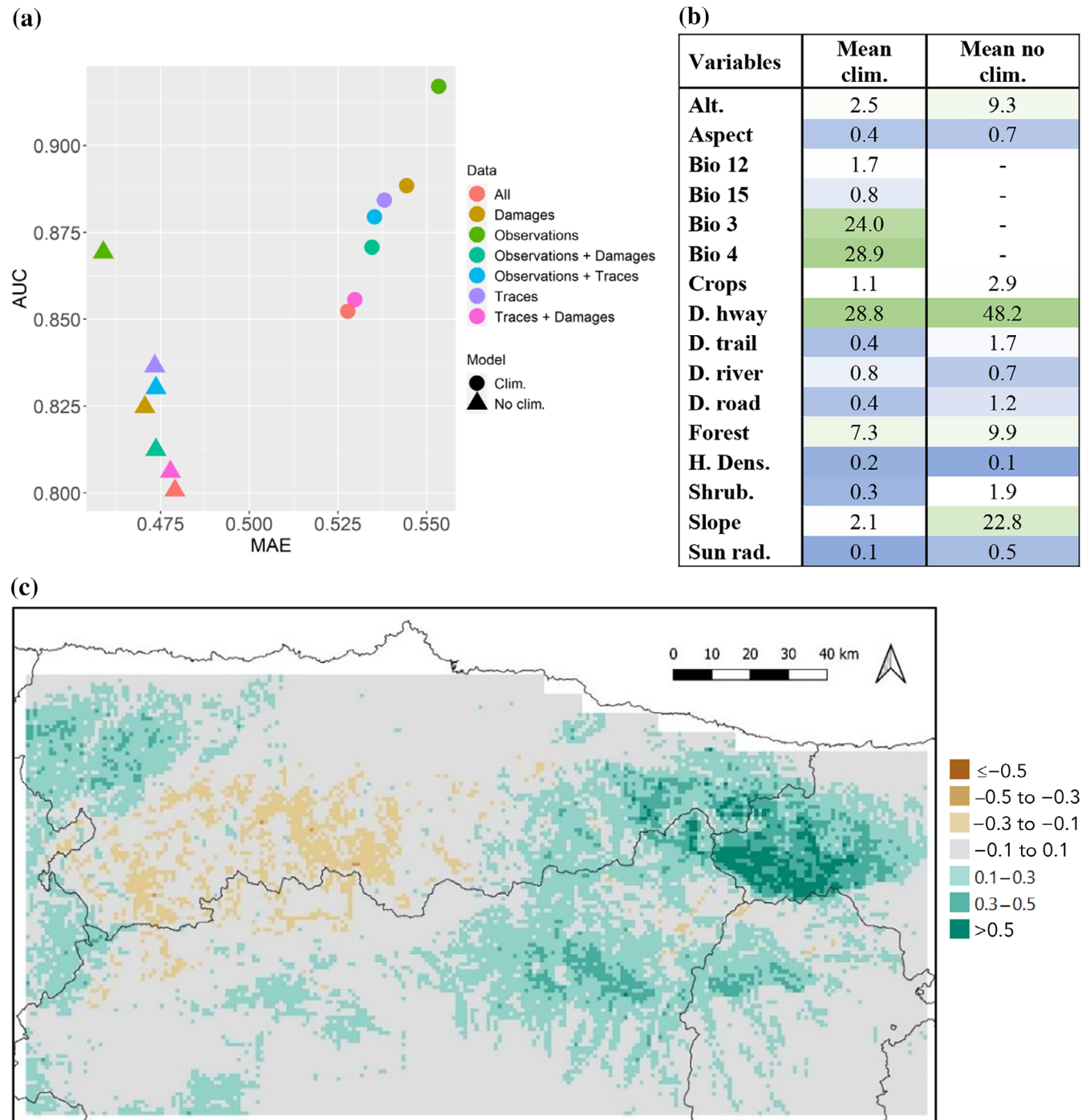lso presented a high mean percentage contribution (54.5%), especially isothermality (24%, Bio 3) and temperature seasonality (29%, Bio 4; Figure 2b). These variables, together with the distance to highways (29%), represent the main drivers of bear habitat suitability in these models. Once climatic variables were removed, the distance to highways increased in importance (+19/48.2% contribution), as well as slope (+20/22.8%) and altitude (+7/9.3%), which, together with forests (9.9%), became the most important explanatory variables. Numerical evaluation of the map predictions of both sets of models showed high similarity (mean Schoener's $I = 0.933$; Appendix S3: Table S1).

On the other hand, models including climatic variables show more restricted habitat suitability and high importance of climatic factors for bear distribution. Taking a closer look at the suitability prediction, the results of climate variables containing models are misleading; this is especially true for the eastern area, where models indicate some of the areas of Cantabria (Figure 2c) to be unsuitable for the bear, while in reality, bears have inhabited these areas for decades (Clevenger et al., 1987). Furthermore, the global brown bear distribution implies good adaptability to a wide variety of climatic conditions (Kaczensky et al., 2013). Therefore, we decided to remove climatic variables from further analysis as they were not appropriate for modeling this bear population's distribution.

Analysis grain and type of occurrence

Models performed better at finer analysis grains, especially in terms of AUC and MAE (0.11 AUC and 0.014 MAE mean difference between the coarsest and finest analysis grain) (Figure 3a), although only the 5-km analysis grain slightly underperformed the other analysis grains according to the Boyce index (Appendix S2: Table S2). At this analysis grain, models constructed using different sets of input variables were mutually more similar than at the other grains (mean Schoener's $I = 0.994$; Appendix S3: Table S2). On the other hand, the importance of individual variables varied more among models at finer grains (Figure 3b). Distance to highways was the most important variable at all analysis grains, while crops and altitude stand out at the 5-km grain, and slope and forest at 1-km and 500-m grains.

According to validation values, we can conclude that all models performed properly at all analysis grains for the different types of data and their combinations (AUC > 0.7, Figure 3a; Boyce index ≈ 1, Appendix S2: Table S2). Considering both the highest AUC and lowest MAE mean values of all combinations of data (Appendix S2: Table S1), models including only

(a)

(b)

Variables	Mean clim.	Mean no clim.
Alt.	2.5	9.3
Aspect	0.4	0.7
Bio 12	1.7	-
Bio 15	0.8	-
Bio 3	24.0	-
Bio 4	28.9	-
Crops	1.1	2.9
D. hway	28.8	48.2
D. trail	0.4	1.7
D. river	0.8	0.7
D. road	0.4	1.2
Forest	7.3	9.9
H. Dens.	0.2	0.1
Shrub.	0.3	1.9
Slope	2.1	22.8
Sun rad.	0.1	0.5

(c)

FIGURE 2 Comparison of the preliminary 1-km models with versus without climatic (clim.) variables according to (a) area under the receiver-operator curve (AUC) and mean absolute error (MAE) values, (b) variable percentage contribution, and (c) suitability map predictions difference; higher predicted habitat suitability is shown in brown for the model not including climatic variables, and in blue for the model containing them, while gray areas show very similar suitability predicted by both for the mean of all types of brown bear data (observations, traces, damages, and their combinations). Alt., altitude; Bio 3, isothermality; Bio 4, temperature seasonality; Bio 12, annual precipitation; Bio 15, precipitation seasonality; D. hway, distance to highway; D. river, distance to river; D. road, distance to road; D. trail, distance to trail; H. Dens., human population density; Shrub., shrubland; Sun rad., sun radiation.

observations appear to perform best (Figure 3a), although the difference in the performance values is very low (0.06 AUC and 0.015 MAE maximum difference). All types of data also showed very similar results in terms of

map predictions (mean Schoener's $I = 0.985$). Models present slight changes in contributions of the most important variables, particularly between analysis grains (Figure 3b). Models based on observations only and

(a)
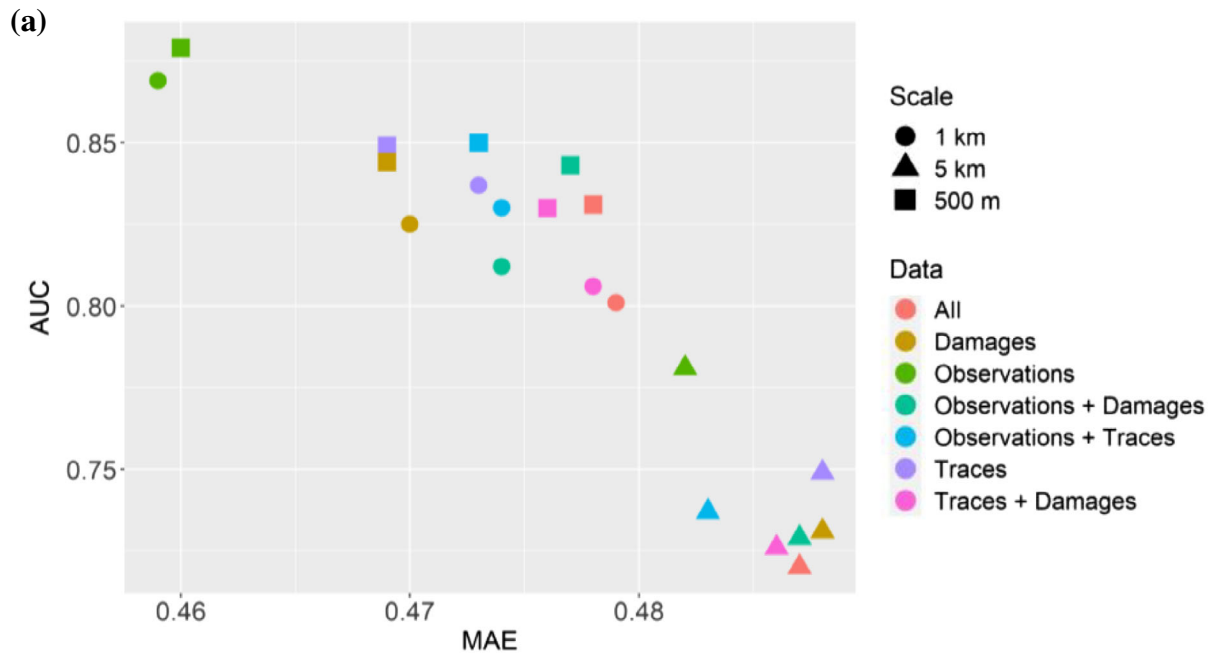
(b)

Scale	Type of data	Alt.	Aspect	Crops	D. hway	D. trail	D. river	D. road	Forest	H. dens.	Shrub.	Slope	Sun rad.
5 km	**Observations**	25	0	21	26	2	0	0	7	0	6	12	0
	Traces	6	0	33	48	0	0	0	2	0	1	8	0
	Damages	5	0	37	48	3	0	0	3	0	2	1	0
	Observations + Damages	19	0	28	35	2	0	0	5	0	6	4	0
	Observations + Traces	19	5	26	22	0	0	0	2	0	4	10	12
	Traces + Damages	13	3	33	32	1	0	0	3	0	2	5	8
	All	12	9	29	26	1	0	0	3	0	2	3	16
1 km	**Observations**	14	1	0	42	1	0	2	11	0	1	27	1
	Traces	10	1	3	46	0	0	0	12	0	1	26	1
	Damages	7	1	5	46	7	4	3	7	1	2	16	0
	Observations + Damages	5	1	2	53	1	0	2	9	0	1	23	1
	Observations + Traces	11	0	2	46	0	0	0	11	0	3	26	1
	Traces + Damages	8	0	5	51	1	0	0	10	0	3	20	0
	All	9	0	3	53	0	0	1	9	0	2	22	0
500 m	**Observations**	9	0	1	53	1	0	2	8	0	0	25	1
	Traces	10	0	2	48	2	0	1	14	0	0	24	0
	Damages	9	1	5	44	10	5	4	5	1	2	14	1
	Observations + Damages	6	1	1	50	5	2	4	7	1	2	21	2
	Observations + Traces	8	1	1	50	2	0	1	12	0	0	25	1
	Traces + Damages	7	1	2	51	3	1	2	11	1	1	20	0
	All	8	0	1	54	2	1	2	10	0	0	21	1

FIGURE 3 Comparison of models for all types of data (observations, traces, damages, and their combinations) according to (a) area under the receiver-operator curve (AUC) and mean absolute error (MAE) values, (b) variable percentage contribution (decreasing from green to blue), and (c) suitability map predictions for the best performing type of data (observations) compared to the best model in terms of ecology (observations and traces) at the finest analysis grain (500 m), where more differences can be observed; higher predicted habitat suitability is shown in brown colors for the model built with observations only, and in blues for the model with observations and traces, while gray areas show very similar suitability predicted by both. Alt., altitude; D. hway, distance to highway; D. river, distance to river; D. road, distance to road; D. trail, distance to trail; H. dens., human population density; Shrub., shrubland; Sun rad., sun radiation.
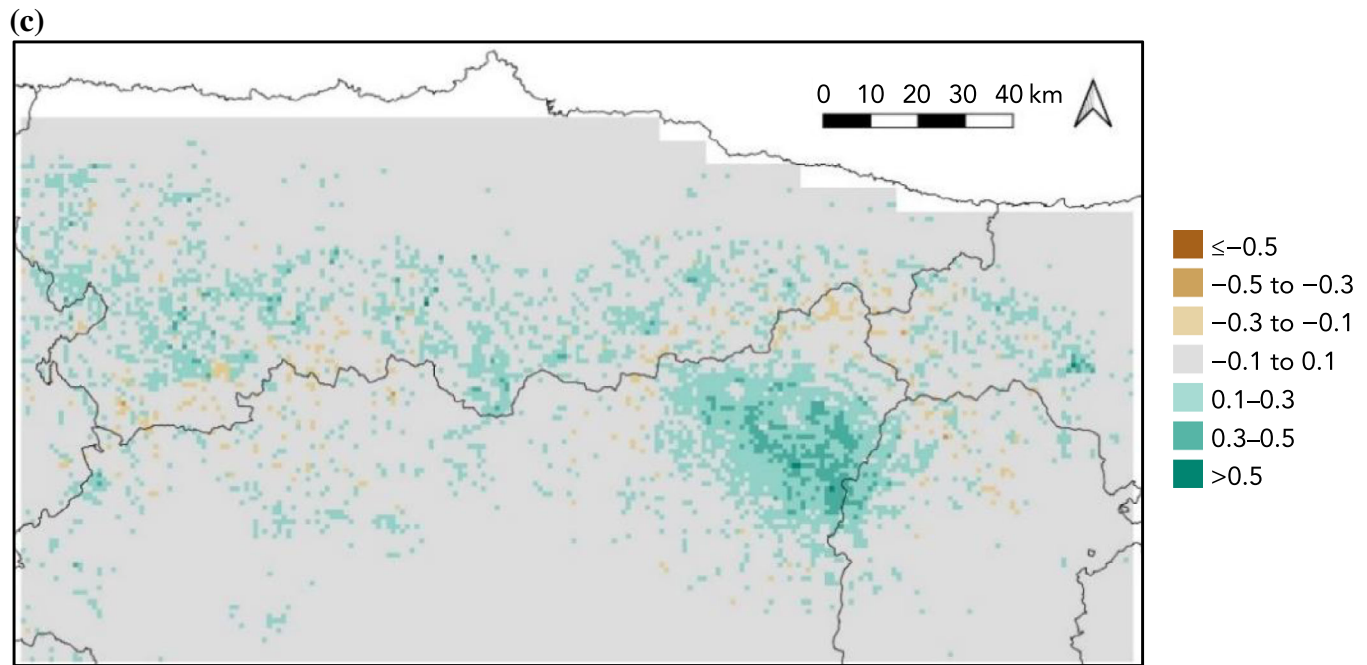
(c)

Legend:
- ≤−0.5
- −0.5 to −0.3
- −0.3 to −0.1
- −0.1 to 0.1
- 0.1–0.3
- 0.3–0.5
- >0.5

FIGURE 3 (Continued)

damages only yielded results more different from the others (regarding variable contribution and Schoener's *I*) (Figure 3b; Appendix S3: Table S2). The combination of observations and traces represents the best presence data for SDMs construction, based on expert knowledge, and gives better predictions (Figure 3c; Appendix S4) than the models performing best in numerical analyses. The latter, including only observations, presented more restricted habitat suitability and underpredicted the eastern subpopulation habitat.
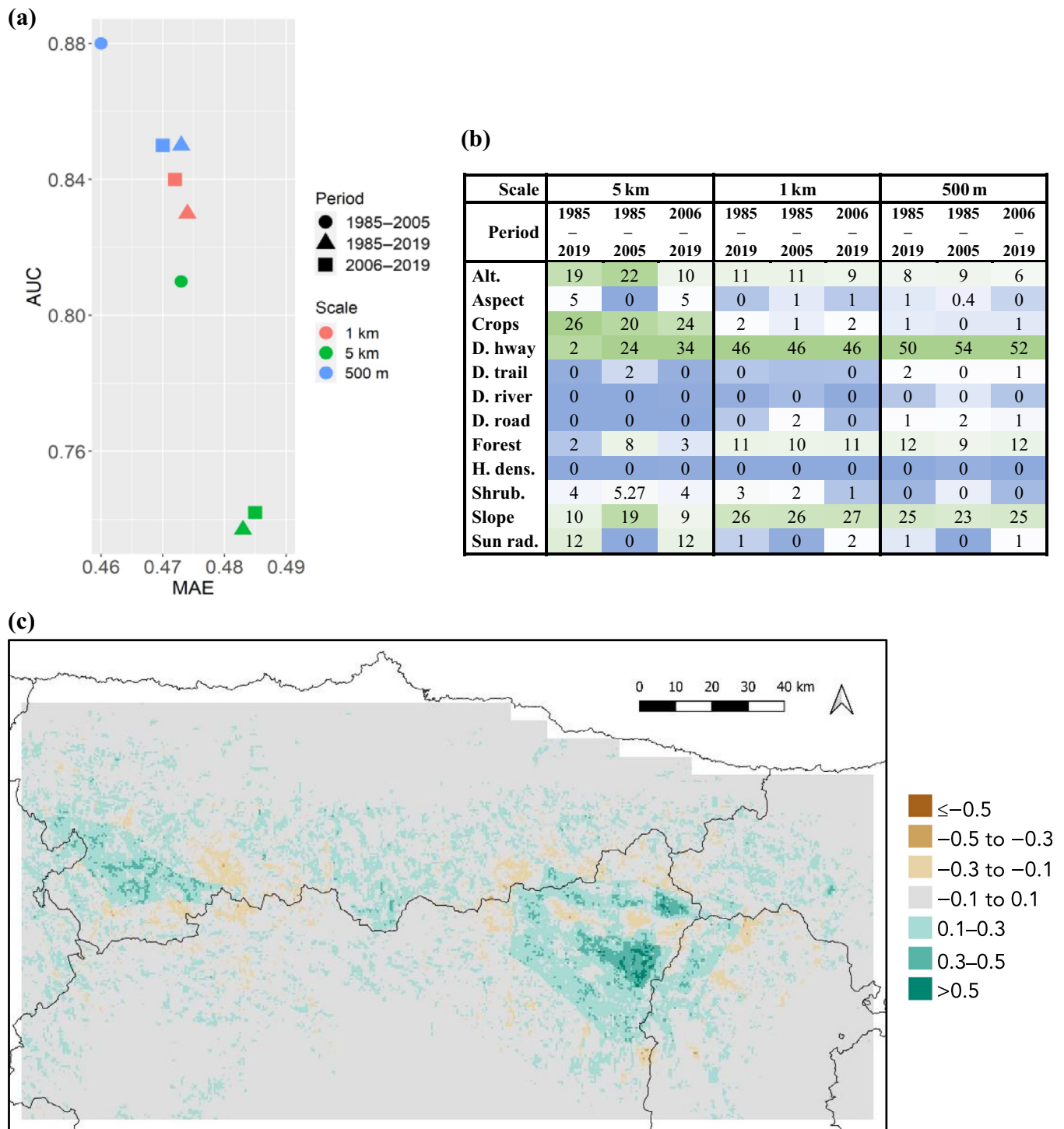
Time frame

Regarding the time frame, all periods produce well-performing models when including both observations and traces at all analysis grains (AUC > 0.74, Figure 4a; Boyce index ≈ 1, Appendix S2: Table S3). Models with data from the first period (1985–2006) stand out, but validation statistics between the entire time frame and the most recent period presented very similar values (0.005 AUC and 0.003 MAE maximum difference, Figure 4a; Appendix S2: Table S3). Both periods also show very similar results in terms of map predictions (Schoener's *I* > 0.99), especially at the larger analysis grains (Appendix S3: Table S3). In the first period, the contribution of individual variables was slightly different from the second period and from the entire time span, especially for the coarsest analysis grain (Figure 4b). By comparing map outputs between periods at the finest analysis grain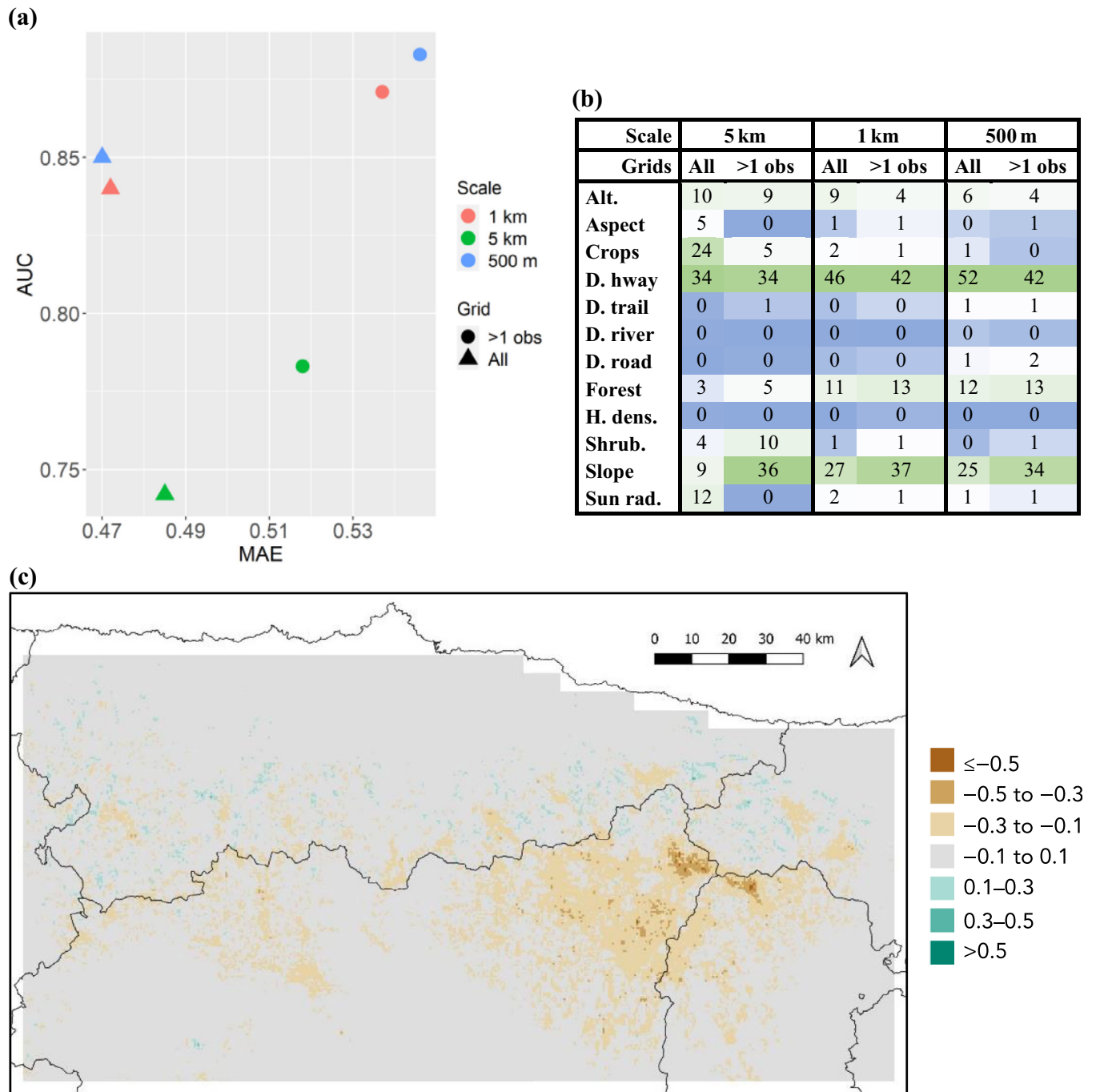 (Figure 4c), predictions do not show evident differences in any specific area, but models from the first period appear more restrictive in terms of habitat suitability (Appendix S5).

Dispersal

When testing the inclusion or exclusion of grids with dispersing bear occurrences, both approaches produce well-performing models at all analysis grains (AUC > 0.74, Figure 5a; Boyce index ≈ 1, Appendix S2: Table S4). However, models considering all grids with more than one occurrence performed better in terms of AUC, although they had a higher MAE. Still, as above, the difference in validation statistics is very low (0.05 and 0.08 units of max AUC and MAE difference between models, respectively; Figure 5a; Appendix S2: Table S3). Map predictions from both models were very similar (Schoener's *I* > 0.98): the coarser the grain, the greater the similarity (Appendix S3: Table S4). In terms of the contribution of individual variables to the models, both models present very similar values, except for the largest analysis grain, where the importance of slope increases at the expense of crops when excluding dispersal grids (Figure 5b). By comparing map outputs from both models at the 500-m analysis grain (which differed most from the other grains), no under- or overprediction in any specific area can be seen. Nevertheless, the model excluding dispersal grids provides a more restricted habitat suitability, avoiding the overestimation of habitat quality in areas used to travel between best landscape patches (Figure 5c; Appendix S6).

(a)

(b)

Scale	5 km			1 km			500 m		
Period	1985–2019	1985–2005	2006–2019	1985–2019	1985–2005	2006–2019	1985–2019	1985–2005	2006–2019
Alt.	19	22	10	11	11	9	8	9	6
Aspect	5	0	5	0	1	1	1	0.4	0
Crops	26	20	24	2	1	2	1	0	1
D. hway	2	24	34	46	46	46	50	54	52
D. trail	0	2	0	0		0	2	0	1
D. river	0	0	0	0	0	0	0	0	0
D. road	0	0	0	0	2	0	1	2	1
Forest	2	8	3	11	10	11	12	9	12
H. dens.	0	0	0	0	0	0	0	0	0
Shrub.	4	5.27	4	3	2	1	0	0	0
Slope	10	19	9	26	26	27	25	23	25
Sun rad.	12	0	12	1	0	2	1	0	1

(c)

FIGURE 4 Comparison of models for the entire time frame (1985–2019) versus both the older (1985–2005) and recent period (2006–2019) using observations and traces as input data according to (a) area under the receiver-operator curve (AUC) and mean absolute error (MAE) values, (b) variable percentage contribution (decreasing from green to blue), and (c) suitability map predictions for the differences between the older and the recent periods at the finest analysis grain (500 m); higher predicted habitat suitability is shown in brown for the model built on the data from the older period, and in blue for the recent period, while gray areas show very similar suitability predicted by both. Alt., altitude; D. hway, distance to highway; D. river, distance to river; D. road, distance to road; D. trail, distance to trail; H. dens., human population density; Shrub., shrubland; Sun rad., sun radiation.
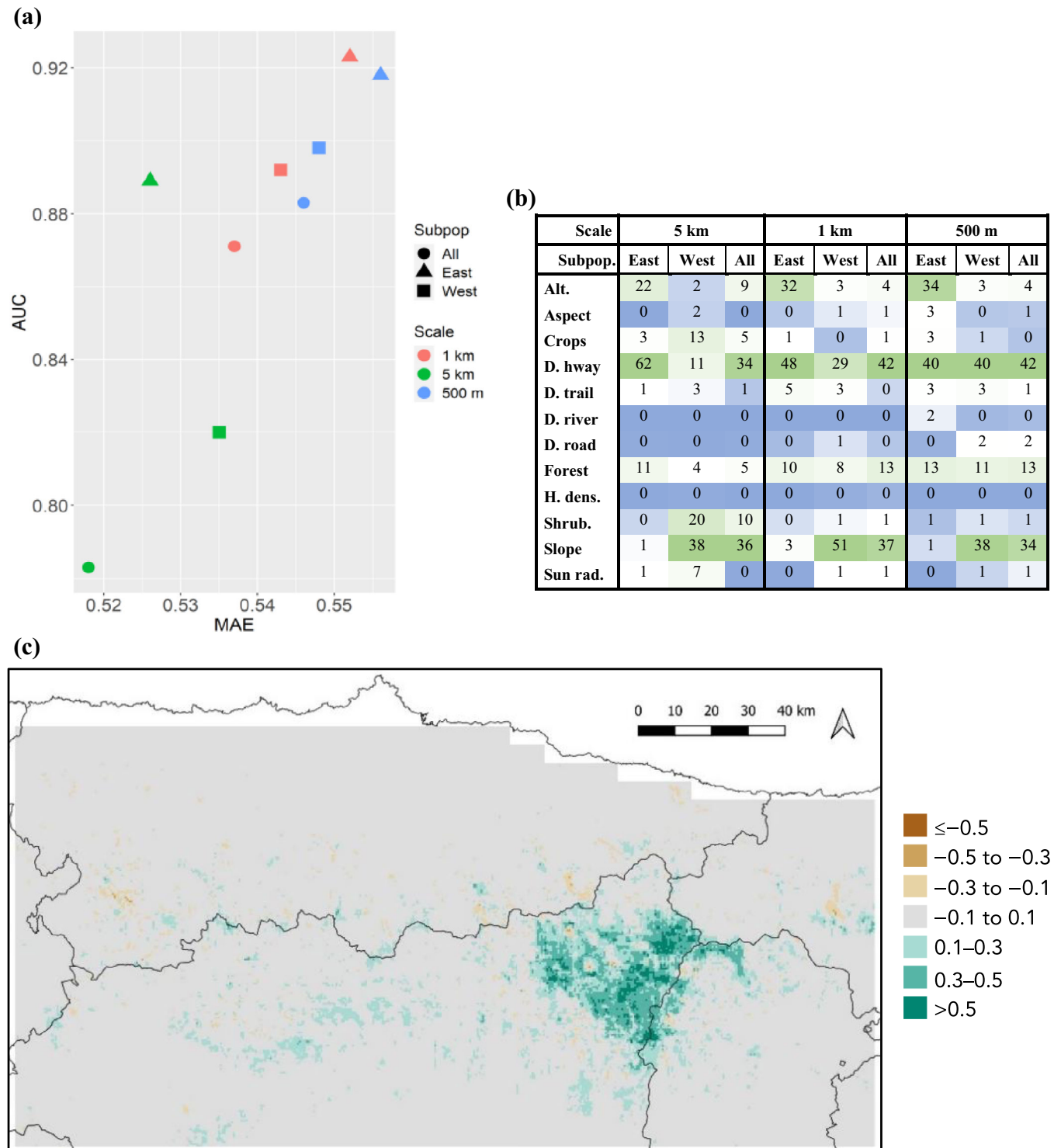
(a)

(b)

Scale	5 km		1 km		500 m	
Grids	**All**	**>1 obs**	**All**	**>1 obs**	**All**	**>1 obs**
Alt.	10	9	9	4	6	4
Aspect	5	0	1	1	0	1
Crops	24	5	2	1	1	0
D. hway	34	34	46	42	52	42
D. trail	0	1	0	0	1	1
D. river	0	0	0	0	0	0
D. road	0	0	0	0	1	2
Forest	3	5	11	13	12	13
H. dens.	0	0	0	0	0	0
Shrub.	4	10	1	1	0	1
Slope	9	36	27	37	25	34
Sun rad.	12	0	2	1	1	1

(c)

FIGURE 5 Comparison of models built with all grids with at least one occurrence (i.e., including dispersal grids) versus grids with more than one occurrence using data on observations (obs) and traces from the period 2006–2019 as input data according to (a) area under the receiver-operator curve (AUC) and mean absolute error (MAE) values, (b) variable percentage contribution (decreasing from green to blue), and (c) suitability map predictions for both periods at the finest analysis grain (500 m); higher predicted habitat suitability is shown in brown for the model including dispersal grids, and in blue for the model excluding them, while gray areas show very similar suitability predicted by both. Alt., altitude; D. hway, distance to highway; D. river, distance to river; D. road, distance to road; D. trail, distance to trail; H. dens., human population density; Shrub., shrubland; Sun rad., sun radiation.

Subpopulations

According to model validation, models including data for the whole population and each subpopulation independently performed well at all analysis grains (AUC > 0.78, Figure 6a; Boyce index ≈ 1, Appendix S2: Table S5). Difference in MAE was very low between models (0.015 unit of maximum MAE difference), but in

(a)

(b)

Scale	5 km			1 km			500 m		
Subpop.	East	West	All	East	West	All	East	West	All
Alt.	22	2	9	32	3	4	34	3	4
Aspect	0	2	0	0	1	1	3	0	1
Crops	3	13	5	1	0	1	3	1	0
D. hway	62	11	34	48	29	42	40	40	42
D. trail	1	3	1	5	3	0	3	3	1
D. river	0	0	0	0	0	0	2	0	0
D. road	0	0	0	0	1	0	0	2	2
Forest	11	4	5	10	8	13	13	11	13
H. dens.	0	0	0	0	0	0	0	0	0
Shrub.	0	20	10	0	1	1	1	1	1
Slope	1	38	36	3	51	37	1	38	34
Sun rad.	1	7	0	0	1	1	0	1	1

(c)

FIGURE 6 Comparison of models constructed using observations and traces from the period of 2006–2019 in all grids with more than one occurrence for the whole population versus the ensemble of models for the western and the eastern subpopulations (Subpop.) at each analysis grain as input data according to (a) area under the receiver-operator curve (AUC) and mean absolute error (MAE) values, (b) variable percentage contribution (decreasing from green to blue), and (c) suitability map predictions for both periods at the finest analysis grain (500 m); higher predicted habitat suitability is shown in brown colors for the model with the entire population, and in blues for the ensemble of the two subpopulations, while gray areas show very similar suitability predicted by both. Alt., altitude; D. hway, distance to highway; D. river, distance to river; D. road, distance to road; D. trail, distance to trail; H. dens., human population density; Shrub., shrubland; Sun rad., sun radiation.

terms of AUC modeling, each subpopulation independently produced better-performing models, especially at the largest analysis grain (Figure 6a). In terms of map predictions, models using the whole population and the combination of the individual subpopulations yielded similar results (mean Schoener's $I = 0.986$), especially at larger analysis grains. They were, however, similar to the models produced when using only the western subpopulation occurrences (Appendix S3: Table S5). Distance to the highway was a highly important variable in all models, especially at the finer analysis grains. The contribution of individual variables, however, differed between subpopulations, with altitude being more important in the eastern subpopulation and slope in the western one, bringing to light that the habitat availability differs between these areas. The contribution of individual variables to the model was similar for the model built on the whole population and the one built for the western population, especially at smaller analysis grains (Figure 6b). This can also be inferred from the map predictions, where the model for the whole population underpredicts the eastern core's habitat suitability (Figure 6c; Appendix S7).

Final comparison

Both the unfiltered model (built once the appropriate variables were selected, i.e., not including climatic variables) and the best model (using observations and traces in 2006–2019 in all grids with more than only one occurrence for the ensemble of independent models of the western and eastern subpopulations) performed well at all analysis grains (AUC > 0.7, Figure 6a; Boyce index ≥0.98, Appendix S2: Tables S1 and S5). Still, the non-filtered model showed lower AUC values (0.134 units of maximum AUC difference in the largest analysis grain) and slightly lower MAE (0.074 units of maximum MAE difference in the smallest analysis grain) than the final chosen model. The two models also showed similar results in terms of map predictions (mean Schoener's $I = 0.977$), especially at the smallest analysis grains (Appendix S3: Table S6). On the other hand, the importance of individual variables varied more. The non-filtered model overestimated the effect of crops and sun radiation at the largest analysis grain, while slope at the largest and altitude at the smallest grains had a low contribution (Figure 7b). Regarding map outputs (Figure 7c), the filtered model gives a more restricted habitat suitability and better reflects the suitable areas in the core of the eastern subpopulation.
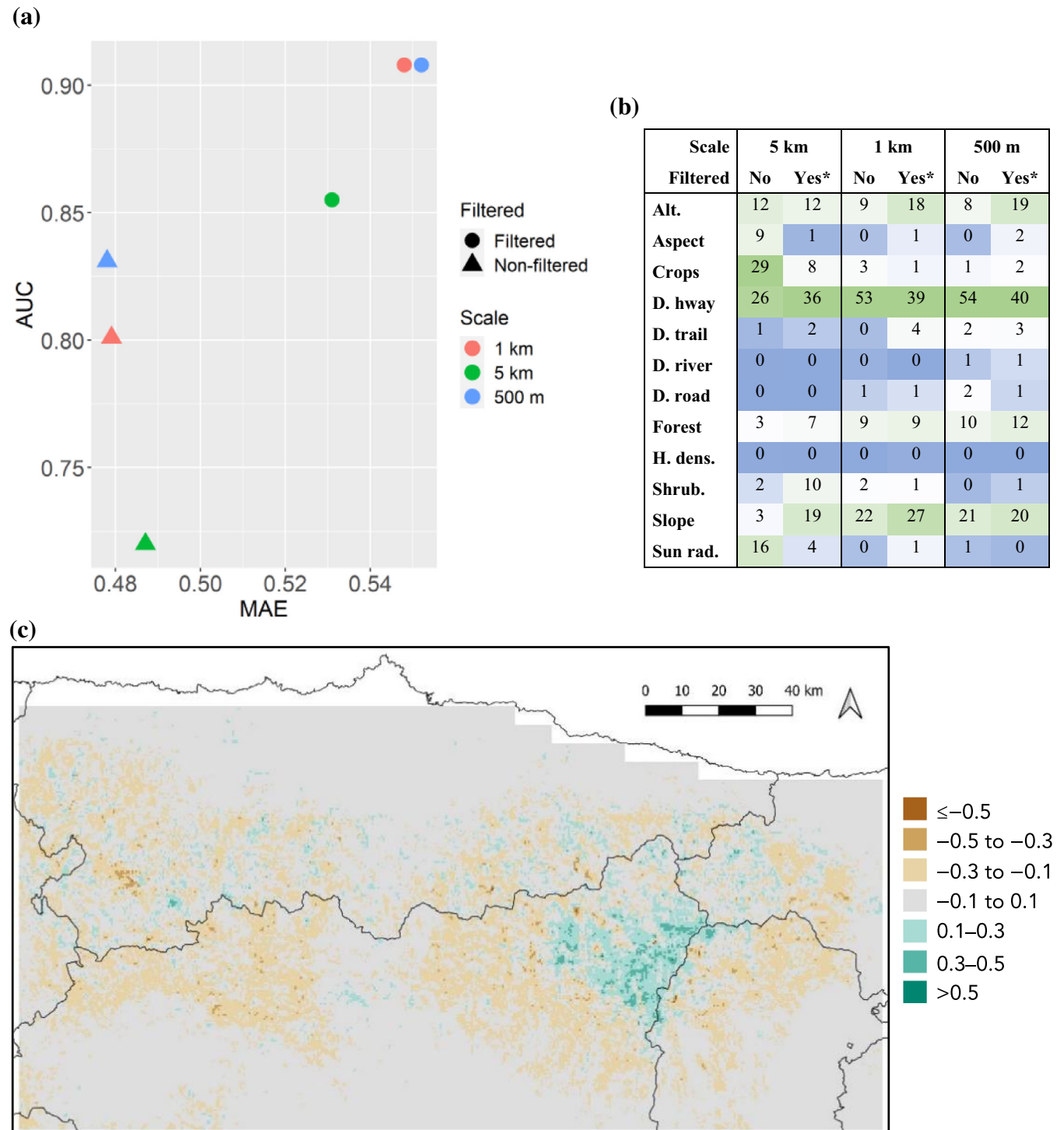
DISCUSSION

SDMs are a useful tool, but some considerations regarding the selection of input data must be addressed before modeling. Several studies have highlighted that before building SDMs, a good ecological and expert background knowledge of the species, its ecology, and distribution are highly recommended (Beck et al., 2014; Harris et al., 2013; Merow et al., 2017). Thus, carefully revising and filtering available data so that they fit the ecological question we are trying to address is necessary to obtain the results in the way most appropriate for their use in developing conservation and management recommendations (Boria et al., 2014; Moudrý & Šímová, 2012). On an example of a generalist species, our results highlight the importance of this, especially when the results are intended for making decisions at the local level.

Climatic variables

Selecting appropriate environmental variables is the key to building proper models. If such models are not ecologically relevant to the species and cannot properly represent conditions that drive species' distribution, results and conclusions can be severely prejudiced (Austin & Van Niel, 2011; Booth, 2021; Gardner et al., 2019). Our results confirm in a real-world case that we cannot trust the models to choose the most important variables if they are not previously "groomed" by expert opinion and knowledge of the species. As demonstrated before (e.g., Harris et al., 2013; Porfirio et al., 2014), the results for the same species can vary greatly depending on the combinations of variables, from predicting almost extinction to very little range contraction. Even though methods for preselecting variables have not yet been explored deeply so far (Williams et al., 2012), some authors have tried to assess the best variable combinations through several methods without much success in finding any universal procedure (Dormann et al., 2008; Maggini et al., 2006; Meynard & Quinn, 2007). It should also be noted that the adequacy of selected variables may also depend on the characteristics of the species, representing a major problem for highly mobile animals (Porfirio et al., 2014). Thus, expert opinion or information about the species should be considered when selecting the environmental drivers.

Furthermore, climatic variables are usually included in these types of studies, but, as evidenced by our models, they may not always yield a good fit. Bears are more likely influenced by climate indirectly, as shown by the fact that they can be found in areas with very different climatic conditions, from Alaska or Russia to Mediterranean

(a)

(b)

Scale	5 km		1 km		500 m	
Filtered	No	Yes*	No	Yes*	No	Yes*
Alt.	12	12	9	18	8	19
Aspect	9	1	0	1	0	2
Crops	29	8	3	1	1	2
D. hway	26	36	53	39	54	40
D. trail	1	2	0	4	2	3
D. river	0	0	0	0	1	1
D. road	0	0	1	1	2	1
Forest	3	7	9	9	10	12
H. dens.	0	0	0	0	0	0
Shrub.	2	10	2	1	0	1
Slope	3	19	22	27	21	20
Sun rad.	16	4	0	1	1	0

(c)

FIGURE 7 Comparison of the non-filtered model (all types of data, i.e., observations, traces, and damages, for the entire time frame [1985–2019] including all grids with at least one occurrence for the whole population) versus the final filtered model (observations and traces in 2006–2019 including all grids with more than only one occurrence for the ensemble of independent models of the western and eastern subpopulations) at each analysis grain according to (a) area under the receiver-operator curve (AUC) and mean absolute error (MAE) values, (b) variable percentage contribution (decreasing from green to blue), and (c) suitability map predictions for both periods at the finest analysis grain (500 m); higher predicted habitat suitability is shown in brown colors for the non-filtered model, and in blues for the final filtered model, while gray areas show very similar suitability predicted by both. *AUC, MAE, and variable percentage contribution values for the filtered model are the means calculated from the independent models of the western and eastern subpopulations. Alt., altitude; D. hway, distance to highway; D. river, distance to river; D. road, distance to road; D. trail, distance to trail; H. dens., human population density; Shrub., shrubland; Sun rad., sun radiation.

countries (and historically even the north of Africa), same as wolves (*Canis lupus*) (Hovardas, 2018; Kaczensky et al., 2013; Ripple et al., 2014). Failure to consider the entire bear distribution (actual but also historical) and, consequently, all climatic conditions this species can endure can lead to misleading results (Chevalier et al., 2021; Faurby & Araújo, 2018). This was also shown in our models that indicated the climatic variables to be the most important ones, but such models led to underprediction of the habitat suitability in areas known to be inhabited by the species. Moreover, as shown by Albrecht et al. (2017), climate (increasing winter temperatures in particular) contributed substantially to the Holocene decline of brown bears mainly indirectly by facilitating human land use. Furthermore, as omnivorous species, bears can adapt to various conditions and food resources depending on natural availability, making these variables even less important (Penteriani et al., 2019; Penteriani & Melletti, 2021; Zarzo-Arias et al., 2020). This emphasizes the need for a basic understanding of the species' ecology for the appropriate selection of variables used for modeling, despite the fact that many studies still overlook this matter. Our findings also imply the urgent need to revise studies on the impact of climate change on these types of species, as the climate itself may fail to predict real shifts in their home range but rather act indirectly by affecting resources (Araújo & Guisan, 2006; Guisan & Thuiller, 2005).

Analysis grain and type of occurrence data

Given that the occurrence data are generally unlikely to be error-free (although the error magnitude can be minimized by, e.g., balanced survey design, even sampling effort, and maintaining low positional uncertainty), we need to focus on how to obtain the right answers by asking the right questions and knowing which data fit our question better (Guillera-Arroita et al., 2015). In our dataset, direct bear observations are usually recorded from the distance in the mountains, either in the clear-cuts in forested parts or above the tree line, as well as from camera traps prevalently located in forested areas where direct observations are more difficult (Penteriani et al., 2019; Zarzo-Arias et al., 2019). Traces include footprints or scats, usually located in already existing trails or mountain paths, and hairs are commonly found in rub trees or fences in the fields (González-Bernardo et al., 2021). On the other hand, damages are linked to human activities, and they are mainly driven by the accessibility of easy resources such as apiaries, fruit trees, or livestock (Zarzo-Arias et al., 2020). Also, they are commonly georeferenced not from the location at which the

damage occurred, but from where the damaged property is registered (i.e., the home of the owner). None of the models with different data combinations stood out in terms of performance, but the importance of individual variables and especially spatial predictions differed, mainly at the largest analysis grain (5 × 5 km). This is of particular concern as bears generally react to the environment at coarser spatial analysis grains (Mateo-Sánchez et al., 2014). Although observations are the species input data type that performed best (Figure 3), map outputs appear misleading as they underestimate the core area of the eastern subpopulation, most likely due to the differences in sampling effort, population size, and habitat characteristics and availability (Lamamy et al., 2019). Thus, ecologically speaking, if looking for habitat suitability, the sensible choice would be to combine observations and traces, as those are the occurrences showing the natural behavior of the species. Damages, on the other hand, would not make good input data as animals approach human activities predominantly due to the presence of easily accessible resources, thus not reflecting the natural behavior (Zarzo-Arias et al., 2020).

Time frame

The dynamics of some carnivore populations in Europe have changed in recent years, going from relict and almost extinct to expanding or even recolonizing parts of their historical ranges, from which they were driven (mainly) due to human activities. This is especially supported by conservation and management efforts increasingly focusing on mitigating human impact on populations with poor conservation status (Chapron et al., 2014; Ripple et al., 2014). Indeed, the brown bears' population in the Cantabrian Mountains has also improved from slow recovery to population increase and colonization of new areas (Mateo-Sánchez et al., 2015; Penteriani et al., 2021; Zarzo-Arias et al., 2019). Occurrences from the first years of monitoring show only very specific and limited areas to be suitable. At that time, bears occupied only the most suitable areas because they did not have to compete with other individuals for space due to the low population size and density (Matthiopoulos et al., 2015). However, using data from the most recent period would be the sensible choice for making decisions based on the current status of the population. As it is currently expanding, the bears might, providing that their conservation remains a priority and human-induced mortality is minimum, soon colonize even landscapes that are more human-disturbed or less abundant in terms of resources (Ciucci & Boitani, 2008; Morales-González et al., 2020). Furthermore, the

technology used in the early years of monitoring was less accurate than in the 21st century with novel technologies (Rodríguez et al., 2007), and sampling effort in the eastern subpopulation was lower in the first period (Fernández-Gil et al., 2010), making these data even poorer in quality.

Dispersal

Considering all cells with at least one record as presences can shift habitat selection at the population level, while reviewing and screening species data usually produce better-performing models (Boria et al., 2014). As mentioned before, individuals in increasing and saturated populations might have to compete for the most suitable habitats or try to find new favorable territories further from their actual core/home range. Among bears, young males are the typical dispersal individuals: they are commonly relegated to the less suitable habitats when fighting for the landscape in a density-saturated population (Zedrosser et al., 2007). They can travel throughout unsuitable landscapes when moving toward new favorable habitat patches, which may place them closer to humans even if bears generally avoid them (Martin et al., 2010; Mateo-Sánchez et al., 2014; Stamps, 2001). In fact, at the coarsest grain of our models, the variable "crops" appears as decisive in explaining the bear's habitat suitability (24% contribution) when including dispersal grids (cells with at least one occurrence). In turn, excluding dispersal grids produces models with more restricted habitat suitability, showing avoidance of more human-modified areas (Figure 5c) and demonstrates increased importance of the variable "slope" as bears' home range is usually linked to steeper territories (Martin et al., 2012).

Subpopulations

When using SDMs, we intend to obtain information about the species' habitat suitability, but specific conditions of the landscape inhabited by a single subpopulation might show only a part of the habitat characteristics that may drive our species' habitat selection, mainly due to the landscape available in their specific areas (Arthur et al., 1996; Mysterud & Ims, 1998). Carnivores, for example, have been restricted to less human-disturbed habitats as they tend to avoid human activities, and the habitat availability may vary throughout the landscape (López-Bao et al., 2017). Indeed, models performed better when focusing on individual subpopulations (Figure 6a). The importance of individual variables differed between the subpopulation models, showing favorable habitat features according to intrinsic habitat availability (Lamamy et al., 2019). This has been also reported in other bear populations (e.g., black bear, Benson & Chamberlain, 2007; or polar bears, Ferguson et al., 2000) as well as in other species (e.g., bison, Kuemmerle et al., 2018; or mice, Wright & Frey, 2015), where different subpopulations or clusters show different habitat preferences. Besides, our models show that when modeling the whole population, the size of the population or the sampling effort in each area can shift the results toward the subpopulation with more available information (western subpopulation, Figure 6b,c). Ensembling these models allows the combining of different subpopulations, species, or even model approaches in order to obtain merged predictions, which could be a more appropriate way to model species populations (Araújo & New, 2007).

Conservation applications

In terms of applied conservation, the fact that there are still favorable and unoccupied habitats in the eastern area of the Cantabrian Mountains (Figure 8) represents the most important result obtained in this study. This, again, brings our attention to the fact that the subpopulation inhabiting this area has not experienced the same expansion and recovery as the western population (Penteriani et al., 2018). "Eastern" bears have been even shown to move to the western areas (Gregório et al., 2020) and have slightly (but not significantly) lower reproductive rate and success (Penteriani et al., 2018), which cannot be explained by differences in habitat composition and structure (Lamamy et al., 2019); this part of the region has even more protected areas. As stated before (Lamamy et al., 2019; Zarzo-Arias et al., 2019), the factors driving this situation in the eastern brown bear Cantabrian subpopulation require urgent study to maintain the quality and conservation status of this population and to enhance the connectivity in the future to increase the gene flow.

CONCLUSIONS

On the example of a generalist species, this study offers a detailed insight into methods for tackling specific types of bias that can be found in data available for building SDMs. In particular, we conclude that for a species such as the brown bear, selection of relevant variables is more important than filtering species' occurrences. Since there is practically no background information on how to
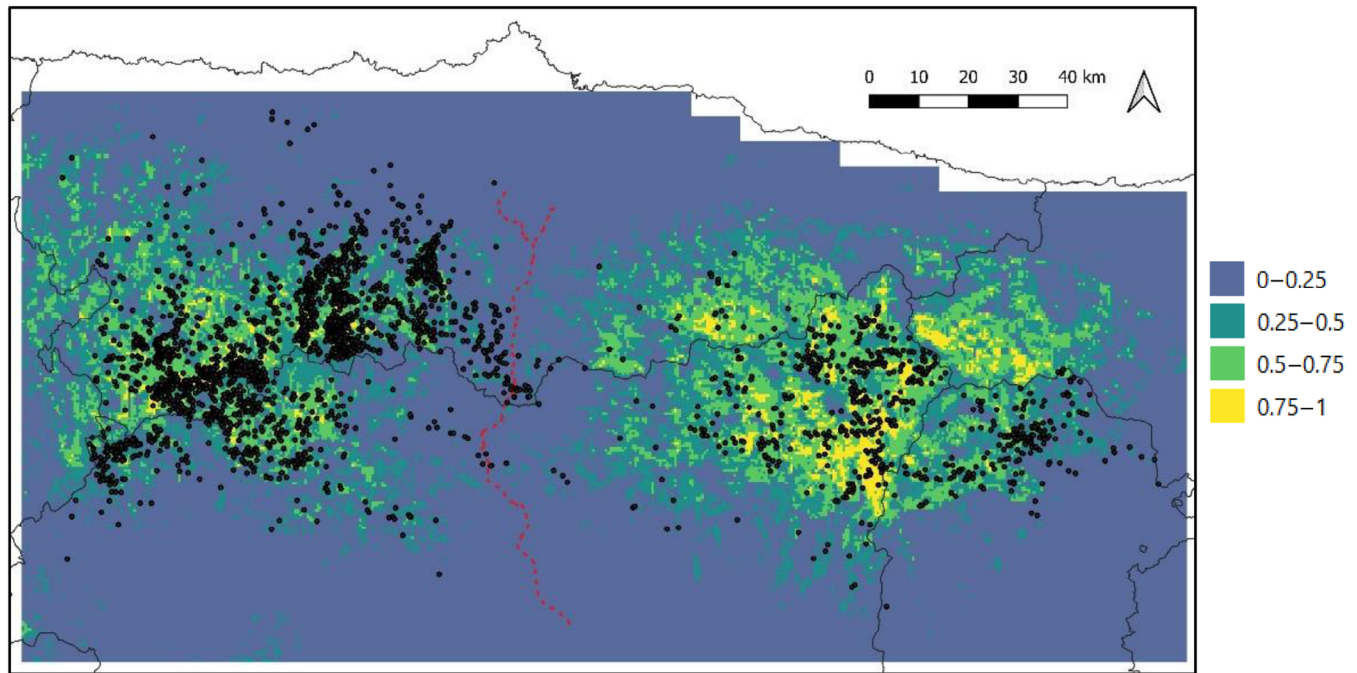
FIGURE 8   Final map prediction without climatic variables, using bear observations and traces data for the most recent period (2006–2019), including all grids with more than only one occurrence and combining models for each subpopulation independently at the finest analysis grain (500 m). Bear occurrences are shown as black points, and the dotted red line indicates the highway dividing the two bear subpopulations.

select appropriate variables for modeling purposes (Dormann et al., 2008; Maggini et al., 2006; Meynard & Quinn, 2007), we show that we cannot expect models themselves to "decide" if the selected variables are indeed a good fit. Furthermore, we show that we cannot just rely on model performance and/or validation metrics to tell us if our model predicts the species' habitat requirements correctly. We first need to choose both the variables (Austin & Van Niel, 2011; Booth, 2021; Gardner et al., 2019) and the occurrence data that are relevant to our research question (Guillera-Arroita et al., 2015) based on the basic knowledge of our species' ecology, in order to obtain appropriate results on which to base applied management and conservation decisions. This is especially important if their application is required at a local level and for species that react to the environment at large spatial scales. Additionally, models can help answer different questions using species input data with different filtering in order to, for example, explore habitat selection by individuals of different age (Milanesi et al., 2016) or sex (Kwon et al., 2019), different life cycle stages (Mateo-Sánchez et al., 2015), or focus only on damages to get conflict hotspots (Behdarvand et al., 2014). In short, we highlight several widely overlooked issues preceding the construction of SDMs that can affect important predictions on species' future and bring the issues to the attention of ecologists

building SDMs. The application of these ideas can improve the selection of appropriate data for building SDMs that address the posed ecological questions and highlight the need for expert knowledge on the target species.

ACKNOWLEDGMENTS

Víctor Vázquez, and Pedro García-Rovés of the Principado de Asturias.

CONFLICT OF INTEREST

DATA AVAILABILITY STATEMENT

ORCID

Alejandra Zarzo-Arias https://orcid.org/0000-0001-5496-0144

Vincenzo Penteriani https://orcid.org/0000-0002-9333-7846

REFERENCES

Albrecht, J., K. A. Bartoń, N. Selva, R. S. Sommer, J. E. Swenson, and R. Bischof. 2017. "Humans and Climate Change Drove the Holocene Decline of the Brown Bear." *Scientific Reports* 7(1): 1–11. https://doi.org/10.1038/s41598-017-10772-6.

Anderson, R. P. 2012. "Harnessing the World's Biodiversity Data: Promise and Peril in Ecological Niche Modeling of Species Distributions." *Annals of the New York Academy of Sciences* 1260(1): 66–80. https://doi.org/10.1111/j.1749-6632.2011.06440.x.

Araújo, M. B., and A. Guisan. 2006. "Five (or So) Challenges for Species Distribution Modelling." *Journal of Biogeography* 33(10): 1677–88. https://doi.org/10.1111/j.1365-2699.2006.01584.x.

Araújo, M. B., and M. New. 2007. "Ensemble Forecasting of Species Distributions." *Trends in Ecology and Evolution* 22(1): 42–7. https://doi.org/10.1016/j.tree.2006.09.010.

Arthur, S. M., B. F. Manly, L. L. McDonald, and G. W. Garner. 1996. "Assessing Habitat Selection When Availability Changes." *Ecology* 77(1): 215–27. https://doi.org/10.2307/2265671.

Austin, M. P., and K. P. Van Niel. 2011. "Improving Species Distribution Models for Climate Change Studies: Variable Selection and Scale." *Journal of Biogeography* 38(1): 1–8. https://doi.org/10.1111/j.1365-2699.2010.02416.x.

Beck, J., M. Böller, A. Erhardt, and W. Schwanghart. 2014. "Spatial Bias in the GBIF Database and Its Effect on Modeling Species' Geographic Distributions." *Ecological Informatics* 19: 10–5. https://doi.org/10.1016/j.ecoinf.2013.11.002.

Behdarvand, N., M. Kaboli, M. Ahmadi, E. Nourani, A. S. Mahini, and M. A. Aghbolaghi. 2014. "Spatial Risk Model and Mitigation Implications for Wolf-Human Conflict in a Highly Modified Agroecosystem in Western Iran." *Biological Conservation* 177: 156–64. https://doi.org/10.1016/j.biocon.2014.06.024.

Benson, J. F., and M. J. Chamberlain. 2007. "Space Use and Habitat Selection by Female Louisiana Black Bears in the Tensas River Basin of Louisiana." *Journal of Wildlife Management* 71(1): 117–26. https://doi.org/10.2193/2005-580.

Berteaux, D., and N. C. Stenseth. 2006. "Measuring, Understanding and Projecting the Effects of Large-Scale Climatic Variability on Mammals." *Climate Research* 32(2): 95–7. https://doi.org/10.3354/cr032095.

Bloom, T. D. S., A. Flower, and E. G. DeChaine. 2018. "Why Georeferencing Matters: Introducing a Practical Protocol to Prepare Species Occurrence Records for Spatial Analysis." *Ecology and Evolution* 8(1): 765–77. https://doi.org/10.1002/ece3.3516.

Booth, T. H. 2021. "A Problem with Variable Selection in a Comparison of Correlative and Process-Based Species Distribution Models: Comments on Higgins et al., 2020." *Ecology and Evolution* 11(19): 13609–12. https://doi.org/10.1002/ece3.7496.

Boria, R. A., L. E. Olson, S. M. Goodman, and R. P. Anderson. 2014. "Spatial Filtering to Reduce Sampling Bias Can Improve the Performance of Ecological Niche Models." *Ecological Modelling* 275: 73–7. https://doi.org/10.1016/j.ecolmodel.2013.12.012.

Brotons, L., S. Herrando, and M. Pla. 2007. "Updating Bird Species Distribution at Large Spatial Scales: Applications of Habitat Modelling to Data from Long-Term Monitoring Programs." *Diversity and Distributions* 13(3): 276–88. https://doi.org/10.1111/j.1472-4642.2007.00339.x.

Brun, P., W. Thuiller, Y. Chauvier, L. Pellissier, R. O. Wüest, Z. Wang, and N. E. Zimmermann. 2020. "Model Complexity Affects Species Distribution Projections under Climate Change." *Journal of Biogeography* 47(1): 130–42. https://doi.org/10.1111/jbi.13734.

Burnham, K. P., and D. R. Anderson. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Berlin: Springer-Verlag.

Chapron, G., P. Kaczensky, J. D. C. Linnell, M. von Arx, D. Huber, H. Andrén, J. V. López-Bao, et al. 2014. "Recovery of Large Carnivores in Europe's Modern Human-Dominated Landscapes." *Science* 346(6216): 1517–9. https://doi.org/10.1126/science.1257553.

Chevalier, M., O. Broennimann, J. Cornuault, and A. Guisan. 2021. "Data Integration Methods to Account for Spatial Niche Truncation Effects in Regional Projections of Species Distribution." *Ecological Applications* 31(7): e02427. https://doi.org/10.1002/eap.2427.

Ciucci, P., and L. Boitani. 2008. "The Apennine Brown Bear: A Critical Review of Its Status and Conservation Problems." *Ursus* 19(2): 130–45. https://doi.org/10.2192/07PER012.1.

Clevenger, A. P., F. J. Purroy, and M. S. de Buruaga. 1987. "Status of the Brown Bear in the Cantabrian Mountains, Spain." *Bears: Their Biology and Management* 7: 1–8.

Dormann, C. F., O. Purschke, J. R. García, S. L. Márquez, and B. Schröder. 2008. "Components of Uncertainty in Species Distribution Analysis: A Case Study of the Great Grey

Shrike." *Ecology* 89(12): 3371–86. https://doi.org/10.1890/07-1772.1.

El-Gabbas, A., and C. F. Dormann. 2018. "Improved Species-Occurrence Predictions in Data-Poor Regions: Using Large-Scale Data and Bias Correction with Down-Weighted Poisson Regression and Maxent." *Ecography* 41(7): 1161–72. https://doi.org/10.1111/ecog.03149.

Elith, J., S. J. Phillips, T. Hastie, M. Dudík, Y. E. Chee, and C. J. Yates. 2011. "A Statistical Explanation of MaxEnt for Ecologists." *Diversity and Distributions* 17(1): 43–57. https://doi.org/10.1111/j.1472-4642.2010.00725.x.

Eriksson, T., and F. Dalerum. 2018. "Identifying Areas for an Expanding Wolf Population." *Biological Conservation* 220: 170–81. https://doi.org/10.1016/j.biocon.2018.02.019.

Faurby, S., and M. B. Araújo. 2018. "Anthropogenic Range Contractions Bias Species Climate Change Forecasts." *Nature Climate Change* 8(3): 252–6. https://doi.org/10.1038/s41558-018-0089-x.

Ferguson, S. H., M. K. Taylor, and F. Messier. 2000. "Influence of Sea Ice Dynamics on Habitat Selection by Polar Bears." *Ecology* 81(3): 761–72. https://doi.org/10.1890/0012-9658(2000)081[0761:IOSIDO]2.0.CO;2.

Fernandes, R. F., D. Scherrer, and A. Guisan. 2018. "How Much Should One Sample to Accurately Predict the Distribution of Species Assemblages? A Virtual Community Approach." *Ecological Informatics* 48: 125–34. https://doi.org/10.1016/J.ECOINF.2018.09.002.

Fernández-Gil, A., A. Ordiz, and J. Naves. 2010. "Are Cantabrian Brown Bears Recovering?" *Ursus* 21(1): 121–4. https://doi.org/10.2192/08SC028.1.

Field, R., B. A. Hawkins, H. V. Cornell, D. J. Currie, J. Alexandre, F. Diniz-Filho, J.-F. Guégan, D. M. Kaufman, et al. 2009. "Spatial Species-Richness Gradients across Scales: A Meta-Analysis." *Journal of Biogeography* 36: 132–47. https://doi.org/10.1111/j.1365-2699.2008.01963.x.

Fielding, A. H., and J. F. Bell. 1997. "A Review of Methods for the Assessment of Prediction Errors in Conservation Presence/Absence Models." *Environmental Conservation* 24(1): 38–49. https://doi.org/10.1017/S0376892997000088.

Fourcade, Y., A. G. Besnard, and J. Secondi. 2018. "Paintings Predict the Distribution of Species, or the Challenge of Selecting Environmental Predictors and Evaluation Statistics." *Global Ecology and Biogeography* 27(2): 245–56. https://doi.org/10.1111/geb.12684.

Fourcade, Y., J. O. Engler, D. Rödder, and J. Secondi. 2014. "Mapping Species Distributions with MAXENT Using a Geographically Biased Sample of Presence Data: A Performance Assessment of Methods for Correcting Sampling Bias." *PLoS One* 9(5): 1–13. https://doi.org/10.1371/journal.pone.0097122.

Gábor, L., V. Moudrý, V. Barták, and V. Lecours. 2020. "How Do Species and Data Characteristics Affect Species Distribution Models and When to Use Environmental Filtering?" *International Journal of Geographical Information Science* 34(8): 1567–84. https://doi.org/10.1080/13658816.2019.1615070.

Gábor, L., V. Moudrý, V. Lecours, M. Malavasi, V. Barták, M. Fogl, P. Šímová, D. Rocchini, and T. Václavík. 2020. "The Effect of Positional Error on Fine Scale Species Distribution Models

Increases for Specialist Species." *Ecography* 43(2): 256–69. https://doi.org/10.1111/ecog.04687.

Gábor, L., P. Šímová, P. Keil, A. Zarzo-Arias, C. J. Marsh, D. Rocchini, M. Malavasi, V. Barták, and V. Moudrý. 2022. "Habitats as Predictors in Species Distribution Models: Shall We Use Continuous or Binary Data?" *Ecography* 2022: 1–9. https://doi.org/10.1111/ecog.06022.

Gallego, F. J. 2010. "A Population Density Grid of the European Union." *Population and Environment* 31(6): 460–73. https://doi.org/10.1007/s11111-010-0108-y.

Gardner, A. S., I. M. D. Maclean, and K. J. Gaston. 2019. "Climatic Predictors of Species Distributions Neglect Biophysiologically Meaningful Variables." *Diversity and Distributions* 25(8): 1318–33. https://doi.org/10.1111/ddi.12939.

González-Bernardo, E., C. Bagnasco, G. Bombieri, A. Zarzo-Arias, H. Ruiz-Villar, A. Morales-González, C. Lamamy, et al. 2021. "Rubbing Behavior of European Brown Bears: Factors Affecting Rub Tree Selectivity and Density." *Journal of Mammalogy* 102(2): 468–80. https://doi.org/10.1093/jmammal/gyaa170.

Greenberg, C., R. Perry, K. Franzreb, S. Loeb, D. Daniel Saenz, E. W. Rudolph, et al. 2014. "Climate Change and Wildlife in the Southern United States." In *Climate Change Adaptation and Mitigation Options: A Guide for Natural Resource Managers in Southern Forest Ecosystems*, edited by J. M. Vose and K. D. Klepzig, 379–420. Boca Raton, FL: Taylor and Francis Group. https://doi.org/10.1201/b15613-12.

Gregório, I., T. Barros, D. Pando, J. Morante, C. Fonseca, and E. Ferreira. 2020. "Paths for Colonization or Exodus? New Insights from the Brown Bear (*Ursus arctos*) Population of the Cantabrian Mountains." *PLoS One* 15(1): e0227302. https://doi.org/10.1371/journal.pone.0227302.

Grimmett, L., R. Whitsed, and A. Horta. 2020. "Presence-Only Species Distribution Models Are Sensitive to Sample Prevalence: Evaluating Models Using Spatial Prediction Stability and Accuracy Metrics." *Ecological Modelling* 431: 109194. https://doi.org/10.1016/j.ecolmodel.2020.109194.

Guillera-Arroita, G. 2017. "Modelling of Species Distributions, Range Dynamics and Communities under Imperfect Detection: Advances, Challenges and Opportunities." *Ecography* 40: 281–95. https://doi.org/10.1111/ecog.02445.

Guillera-Arroita, G., J. J. Lahoz-Monfort, J. Elith, A. Gordon, H. Kujala, P. E. Lentini, M. A. Mccarthy, R. Tingley, and B. A. Wintle. 2015. "Is My Species Distribution Model Fit for Purpose? Matching Data and Models to Applications." *Global Ecology and Biogeography* 24(3): 276–92. https://doi.org/10.1111/geb.12268.

Guisan, A., and W. Thuiller. 2005. "Predicting Species Distribution: Offering More than Simple Habitat Models." *Ecology Letters* 8(9): 993–1009. https://doi.org/10.1111/j.1461-0248.2005.00792.x.

Guisan, A., R. Tingley, J. B. Baumgartner, I. Naujokaitis-Lewis, P. R. Sutcliffe, A. I. T. Tulloch, T. J. Regan, et al. 2013. "Predicting Species Distributions for Conservation Decisions." *Ecology Letters* 16(12): 1424–35. https://doi.org/10.1111/ele.12189.

Harris, R. M. B., L. L. Porfirio, S. Hugh, G. Lee, N. L. Bindoff, B. Mackey, and N. J. Beeton. 2013. "To Be or Not to Be? Variable Selection Can Change the Projected Fate of a

Threatened Species under Future Climate." *Ecological Management and Restoration* 14(3): 230–4. https://doi.org/10.1111/emr.12055.

Hijmans, R. J., S. Phillips, J. R. Leathwick, and J. Elith. 2017. "Package 'Dismo'." *Circles* 9(1): 1–68.

Hirzel, A. H., G. Le Lay, V. Helfer, C. Randin, and A. Guisan. 2006. "Evaluating the Ability of Habitat Suitability Models to Predict Species Presences." *Ecological Modelling* 199(2): 142–52. https://doi.org/10.1016/j.ecolmodel.2006.05.017.

Hovardas, T. 2018. *Large Carnivore Conservation and Management: Human Dimensions*. London: Routledge. https://doi.org/10.4324/9781315175454.

Huston, M. A. 2002. "Introductory Essay: Critical Issues for Improving Predictions." In *Predicting Species Occurrences: Issues of Accuracy and Scale*, edited by J. M. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson. Covelo, CA: Island Press.

Inman, R., J. Franklin, T. Esque, and K. Nussear. 2021. "Comparing Sample Bias Correction Methods for Species Distribution Modeling Using Virtual Species." *Ecosphere* 12(3): e03422. https://doi.org/10.1002/ecs2.3422.

Isaac, N. J. B., and M. J. O. Pocock. 2015. "Bias and Information in Biological Records." *Biological Journal of the Linnean Society* 115(3): 522–31. https://doi.org/10.1111/bij.12532.

Jayathilake, D. R. M., and M. J. Costello. 2020. "A Modelled Global Distribution of the Kelp Biome." *Biological Conservation* 252: 108815. https://doi.org/10.1016/j.biocon.2020.108815.

Jetz, W., J. M. McPherson, and R. P. Guralnick. 2012. "Integrating Biodiversity Distribution Knowledge: Toward a Global Map of Life." *Trends in Ecology and Evolution* 27(3): 151–9. https://doi.org/10.1016/j.tree.2011.09.007.

Jiménez, L., and J. Soberón. 2020. "Leaving the Area under the Receiving Operating Characteristic Curve behind: An Evaluation Method for Species Distribution Modelling Applications Based on Presence-Only Data." *Methods in Ecology and Evolution* 11(12): 1571–86. https://doi.org/10.1111/2041-210X.13479.

Kaczensky, P., G. Chapron, M. von Arx, D. Huber, H. Andrén, and J. Linnell. 2013. "Status, Management and Distribution of Large Carnivores—Bear, Lynx, Wolf and Wolverine—in Europe." Report to the EU Commission, 272.

Konowalik, K., and A. Nosol. 2021. "Evaluation Metrics and Validation of Presence-Only Species Distribution Models Based on Distributional Maps with Varying Coverage." *Scientific Reports* 11(1): 1–15. https://doi.org/10.1038/s41598-020-80062-1.

Kramer-Schadt, S., J. Niedballa, J. D. Pilgrim, B. Schröder, J. Lindenborn, V. Reinfelder, M. Stillfried, et al. 2013. "The Importance of Correcting for Sampling Bias in MaxEnt Species Distribution Models." *Diversity and Distributions* 19(11): 1366–79. https://doi.org/10.1111/ddi.12096.

Krofel, M., M. Elfström, H. Ambarlı, G. Bombieri, E. González-Bernardo, K. Jerina, A. Laguna, et al. 2021. "Human–Bear Conflicts at the Beginning of the Twenty-First Century: Patterns, Determinants, and Mitigation Measures." In *Bears of the World*, edited by V. Penteriani and M. Melleti, 213–26. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108692571.016.

Kuemmerle, T., C. Levers, B. Bleyhl, W. Olech, K. Perzanowski, C. Reusch, and S. Kramer-Schadt. 2018. "One Size Does Not Fit All: European Bison Habitat Selection across Herds and Spatial Scales." *Landscape Ecology* 33(9): 1559–72. https://doi.org/10.1007/s10980-018-0684-2.

Kuralt, Ž., and I. Kos. 2018. "Open Access Data and Species Distribution Modelling: A Case Study of *Lithobius erythrocephalus* (Chilopoda: Lithobiidae)." *Acta Societatis Zoologicae Bohemicae* 82: 89–96. https://www.researchgate.net/publication/337837200.

Kwon, Y., A. C. Doty, M. L. Huffman, V. Rolland, D. R. Istvanko, and T. S. Risch. 2019. "Implications of Forest Management Practices for Sex-Specific Habitat Use by *Nycticeius humeralis*." *Journal of Mammalogy* 100(4): 1263–73. https://doi.org/10.1093/jmammal/gyz088.

Lamamy, C., G. Bombieri, A. Zarzo-Arias, E. González-Bernardo, and V. Penteriani. 2019. "Can Landscape Characteristics Help Explain the Different Trends of Cantabrian Brown Bear Subpopulations?" *Mammal Research* 64: 559–67. https://doi.org/10.1007/s13364-019-00440-7.

Lecours, V., R. Devillers, A. E. Simms, V. L. Lucieer, and C. J. Brown. 2017. "Towards a Framework for Terrain Attribute Selection in Environmental Studies." *Environmental Modelling and Software* 89: 19–30. https://doi.org/10.1016/j.envsoft.2016.11.027.

Lecours, V., L. Gábor, E. Edinger, and R. Devillers. 2020. "Fine-Scale Habitat Characterization of the Gully, the Flemish Cap, and the Orphan Knoll, Northwest Atlantic, with a Focus on Cold-Water Corals." In *Seafloor Geomorphology as Benthic Habitat* 735–51. Amsterdam: Elsevier. https://doi.org/10.1016/B978-0-12-814960-7.00044-0.

Leitão, P. J., F. Moreira, and P. E. Osborne. 2011. "Effects of Geographical Data Sampling Bias on Habitat Models of Species Distributions: A Case Study with Steppe Birds in Southern Portugal." *International Journal of Geographical Information Science* 25(3): 439–54. https://doi.org/10.1080/13658816.2010.531020.

Lobo, J. M., A. Jiménez-valverde, and R. Real. 2008. "AUC: A Misleading Measure of the Performance of Predictive Distribution Models." *Global Ecology and Biogeography* 17(2): 145–51. https://doi.org/10.1111/j.1466-8238.2007.00358.x.

López-Bao, J. V., J. Bruskotter, and G. Chapron. 2017. "Finding Space for Large Carnivores." *Nature Ecology and Evolution* 1(5): 0140. https://doi.org/10.1038/s41559-017-0140.

Maggini, R., A. Lehmann, N. E. Zimmermann, and A. Guisan. 2006. "Improving Generalized Regression Analysis for the Spatial Prediction of Forest Communities." *Journal of Biogeography* 33(10): 1729–49. https://doi.org/10.1111/j.1365-2699.2006.01465.x.

Marcer, A., A. D. Chapman, J. R. Wieczorek, F. X. Picó, F. Uribe, J. Waller, and A. H. Ariño. 2022. "Uncertainty Matters: Ascertaining Where Specimens in Natural History Collections Come from and Its Implications for Predicting Species Distributions." *Ecography* 2022: e06025. https://doi.org/10.1111/ecog.06025.

Martin, J., M. Basille, B. van Moorter, J. Kindberg, D. Allainé, and J. E. Swenson. 2010. "Coping with Human Disturbance: Spatial and Temporal Tactics of the Brown Bear (*Ursus*

arctos).” *Canadian Journal of Zoology* 88(9): 875–83. https://doi.org/10.1139/Z10-053.

Martin, J., E. Revilla, P.-y. Quenette, J. Naves, J. E. Swenson, and C. Ame. 2012. “Brown Bear Habitat Suitability in the Pyrenees: Transferability across Sites and Linking Scales to Make the Most of Scarce Data.” *Journal of Applied Ecology* 49(3): 621–31. https://doi.org/10.1111/j.1365-2664.2012.02139.x.

Mateo-Sánchez, M. C., S. A. Cushman, and S. Saura. 2014. “Scale Dependence in Habitat Selection: The Case of the Endangered Brown Bear (*Ursus arctos*) in the Cantabrian Range (NW Spain).” *International Journal of Geographical Information Science* 28(8): 1531–46. https://doi.org/10.1080/13658816.2013.776684.

Mateo-Sánchez, M. C., A. Gastón, C. Ciudad, J. I. García-Viñas, J. Cuevas, C. López-Leiva, A. Fernández-landa, et al. 2015. “Seasonal and Temporal Changes in Species Use of the Landscape: How Do They Impact the Inferences from Multi-Scale Habitat Modeling?” *Landscape Ecology* 31(6): 1261–76. https://doi.org/10.1007/s10980-015-0324-z.

Matthiopoulos, J., J. Fieberg, G. Aarts, H. L. Beyer, J. M. Morales, and D. T. Haydon. 2015. “Establishing the Link between Habitat Selection and Animal Population Dynamics.” *Ecological Monographs* 85(3): 413–36. https://doi.org/10.1890/14-2244.1.

Merow, C., M. J. Smith, and J. A. Silander. 2013. “A Practical Guide to MaxEnt for Modeling Species’ Distributions: What It Does, and Why Inputs and Settings Matter.” *Ecography* 36(10): 1058–69. https://doi.org/10.1111/j.1600-0587.2013.07872.x.

Merow, C., A. M. Wilson, and W. Jetz. 2017. “Integrating Occurrence Data and Expert Maps for Improved Species Range Predictions.” *Global Ecology and Biogeography* 26(2): 243–58. https://doi.org/10.1111/geb.12539.

Meynard, C. N., and J. F. Quinn. 2007. “Predicting Species Distributions: A Critical Comparison of the Most Common Statistical Models Using Artificial Species.” *Journal of Biogeography* 34(8): 1455–69. https://doi.org/10.1111/j.1365-2699.2007.01720.x.

Milanesi, P., L. Giraudo, A. Morand, R. Viterbi, and G. Bogliani. 2016. “Does Habitat Use and Ecological Niche Shift over the Lifespan of Wild Species? Patterns of the Bearded Vulture Population in the Western Alps.” *Ecological Research* 31(2): 229–38. https://doi.org/10.1007/s11284-015-1329-4.

Morales, N. S., I. C. Fernández, and V. Baca-González. 2017. “MaxEnt’s Parameter Configuration and Small Samples: Are We Paying Attention to Recommendations? A Systematic Review.” *PeerJ* 2017(3): 1–16. https://doi.org/10.7717/peerj.3093.

Morales-González, A., H. Ruiz-Villar, A. Ordiz, and V. Penteriani. 2020. “Large Carnivores Living alongside Humans: Brown Bears in Human-Modified Landscapes.” *Global Ecology and Conservation* 22: 1–13. https://doi.org/10.1016/j.gecco.2020.e00937.

Moudrý, V., and R. Devillers. 2020. “Quality and Usability Challenges of Global Marine Biodiversity Databases: An Example for Marine Mammal Data.” *Ecological Informatics* 56: 101051. https://doi.org/10.1016/j.ecoinf.2020.101051.

Moudrý, V., J. Komárek, and P. Šímová. 2017. “Which Breeding Bird Categories Should We Use in Models of Species Distribution?” *Ecological Indicators* 74: 526–9. https://doi.org/10.1016/j.ecolind.2016.11.006.

Moudrý, V., V. Lecours, K. Gdulová, L. Gábor, L. Moudrá, J. Kropáček, and J. Wild. 2018. “On the Use of Global DEMs in Ecological Modelling and the Accuracy of New Bare-Earth DEMs.” *Ecological Modelling* 383: 3–9. https://doi.org/10.1016/j.ecolmodel.2018.05.006.

Moudrý, V., V. Lecours, M. Malavasi, B. Misiuk, L. Gábor, K. Gdulová, P. Šímová, and J. Wild. 2019. “Potential Pitfalls in Rescaling Digital Terrain Model-Derived Attributes for Ecological Studies.” *Ecological Informatics* 54: 100987. https://doi.org/10.1016/j.ecoinf.2019.100987.

Moudrý, V., and P. Šímová. 2012. “Influence of Positional Accuracy, Sample Size and Scale on Modelling Species Distributions: A Review.” *International Journal of Geographical Information Science* 26(11): 2083–95. https://doi.org/10.1080/13658816.2012.721553.

Muscarella, R., P. J. Galante, M. Soley-Guardia, R. A. Boria, J. M. Kass, M. Uriarte, and R. P. Anderson. 2014. “ENMeval: An R Package for Conducting Spatially Independent Evaluations and Estimating Optimal Model Complexity for Maxent Ecological Niche Models.” *Methods in Ecology and Evolution* 5(11): 1198–205. https://doi.org/10.1111/2041-210X.12261.

Mysterud, A., and R. A. Ims. 1998. “Functional Responses in Habitat Use: Availability Influences Relative Use in Trade-off Situations.” *Ecology* 79(4): 1435–41. https://doi.org/10.1890/0012-9658(1998)079[1435:FRIHUA]2.0.CO;2.

Newbold, T. 2010. “Applications and Limitations of Museum Data for Conservation and Ecology, with Particular Attention to Species Distribution Models.” *Progress in Physical Geography* 34(1): 3–22. https://doi.org/10.1177/0309133309355630.

Parra, J. L., C. C. Graham, and J. F. Freile. 2004. “Evaluating Alternative Data Sets for Ecological Niche Models of Birds in the Andes.” *Ecography* 27(3): 350–60. https://doi.org/10.1111/j.0906-7590.2004.03822.x.

Pearson, R. G., and T. P. Dawson. 2003. “Predicting the Impacts of Climate Change on the Distribution of Species: Are Bioclimate Envelope Models Useful?” *Global Ecology and Biogeography* 12(5): 361–71. https://doi.org/10.1046/j.1466-822X.2003.00042.x.

Penteriani, V., A. A. Karamanlidis, A. Ordiz, P. Ciucci, L. Boitani, G. Bertorelle, A. Zarzo-Arias, et al. 2021. “Bears in Human-Modified Landscapes: The Case Studies of the Cantabrian, Apennine, and Pindos Mountains.” In *Bears of the World*, edited by V. Penteriani and M. Melleti, 260–72. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108692571.020.

Penteriani, V., and M. Melletti. 2021. In *Bears of the World: Ecology, Conservation and Management*, edited by V. Penteriani and M. Melletti. Cambridge: Cambridge University Press.

Penteriani, V., A. Zarzo-Arias, G. Bombieri, D. Cañedo, J. D. García, M. del Mar Delgado, P. P. Torre, et al. 2018. “Density and Reproductive Characteristics of Female Brown Bears in the Cantabrian Mountains, NW Spain.” *European Zoological Journal* 85(1): 312–20. https://doi.org/10.1080/24750263.2018.1499826.

Penteriani, V., A. Zarzo-Arias, A. Novo-Fernández, G. Bombieri, and C. A. López-Sánchez. 2019. “Responses of an Endangered Brown Bear Population to Climate Change Based on

Predictable Food Resource and Shelter Alterations." *Global Change Biology* 25(3): 1133–51. https://doi.org/10.1111/gcb.14564.

Pérez, T., J. Naves, J. F. Vázquez, A. Fernández-Gil, J. Seijas, J. Albornoz, E. Revilla, M. Delibes, and A. Domínguez. 2014. "Estimating the Population Size of the Endangered Cantabrian Brown Bear through Genetic Sampling." *Wildlife Biology* 20(5): 300–9. https://doi.org/10.2981/wlb.00069.

Phillips, S. J., M. Dudík, and R. E. Schapire. 2004. "A Maximum Entropy Approach to Species Distribution Modeling." In *Proceedings of the Twenty-First International Conference on Machine Learning.* https://doi.org/10.1145/1015330.1015412.

Phillips, S. J., R. T. Anderson, and R. E. Schapire. 2006. "Maximum Entropy Modeling of Species Geographic Distributions." *Ecological Modelling* 190: 231–59. https://doi.org/10.1016/j.ecolmodel.2005.03.026.

Phillips, S. J., and M. Dudík. 2008. "Modeling of Species Distribution with Maxent: New Extensions and a Comprehensive Evaluation." *Ecograpy* 31: 161–75. https://doi.org/10.1111/j.2007.0906-7590.05203.x.

Porfirio, L. L., R. M. B. Harris, E. C. Lefroy, S. Hugh, S. F. Gould, G. Lee, N. L. Bindoff, and B. Mackey. 2014. "Improving the Use of Species Distribution Models in Conservation Planning and Management under Climate Change." *PLoS One* 9(11): 1–21. https://doi.org/10.1371/journal.pone.0113749.

Quantum GIS Development Team. 2015. "QGIS Geographic Information System." Open Source Geospatial Foundation Project. http://qgis.osgeo.org.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing. https://www.R-project.org/.

Rayment, W., S. Dawson, and T. Webster. 2015. "Breeding Status Affects Fine-Scale Habitat Selection of Southern Right Whales on Their Wintering Grounds." *Journal of Biogeography* 42(3): 463–74. https://doi.org/10.1111/jbi.12443.

Renner, I. W., and D. I. Warton. 2013. "Equivalence of MAXENT and Poisson Point Process Models for Species Distribution Modeling in Ecology." *Biometrics* 69(1): 274–81. https://doi.org/10.1111/j.1541-0420.2012.01824.x.

Ripple, W. J., J. A. Estes, R. L. Beschta, C. C. Wilmers, E. G. Ritchie, M. Hebblewhite, J. Berger, et al. 2014. "Status and Ecological Effects of the World's Largest Carnivores." *Science* 343(6167): 1241484. /10.1126/science.1241484.

Rodríguez, J. P., L. Brotons, J. Bustamante, and J. Seoane. 2007. "The Application of Predictive Modelling of Species Distribution to Biodiversity Conservation." *Diversity and Distributions* 13(3): 243–51. https://doi.org/10.1111/j.1472-4642.2007.00356.x.

Seoane, J., and L. M. Carrascal. 2008. "Interspecific Differences in Population Trends of Spanish Birds Are Related to Habitat and Climatic Preferences." *Global Ecology and Biogeography* 17(1): 111–21. https://doi.org/10.1111/j.1466-8238.2007.00351.x.

Stamps, J. A. 2001. "Habitat Selection by Dispersers: Integrating Proximate and Ultimate Approaches." In *Dispersal*, edited by E. D. Clobert, A. Dhondt, and J. Nichols, 230–42. New York: Oxford University Press.

Stockwell, D. R. B., and A. Townsend Peterson. 2002. "Effects of Sample Size on Accuracy of Species Distribution Models."

Ecological Modelling 148(1): 1–13. https://doi.org/10.1016/S0304-3800(01)00388-X.

Swanepoel, L. H., P. Lindsey, M. J. Somers, W. van Hoven, and F. Dalerum. 2013. "Extent and Fragmentation of Suitable Leopard Habitat in South Africa." *Animal Conservation* 16(1): 41–50. https://doi.org/10.1111/j.1469-1795.2012.00566.x.

Tessarolo, G., T. F. Rangel, M. B. Araújo, and J. Hortal. 2014. "Uncertainty Associated with Survey Design in Species Distribution Models." *Diversity and Distributions* 20(11): 1258–69. https://doi.org/10.1111/ddi.12236.

Tomaštík, J., and M. Varga. 2021. "Practical Applicability of Processing Static, Short-Observation-Time Raw GNSS Measurements Provided by a Smartphone under Tree Vegetation." *Measurement: Journal of the International Measurement Confederation* 178: 109397. https://doi.org/10.1016/j.measurement.2021.109397.

Varela, S., R. P. Anderson, R. García-Valdés, and F. Fernández-González. 2014. "Environmental Filters Reduce the Effects of Sampling Bias and Improve Predictions of Ecological Niche Models." *Ecography* 37(11): 1084–91. https://doi.org/10.1111/j.1600-0587.2013.00441.x.

Veloz, S. D. 2009. "Spatially Autocorrelated Sampling Falsely Inflates Measures of Accuracy for Presence-Only Niche Models." *Journal of Biogeography* 36(12): 2290–9. https://doi.org/10.1111/j.1365-2699.2009.02174.x.

Warren, D. L., R. E. Glor, and M. Turelli. 2008. "Environmental Niche Equivalency versus Conservatism: Quantitative Approaches to Niche Evolution." *Evolution: International Journal of Organic Evolution* 62(11): 2868–83. https://doi.org/10.1111/j.1558-5646.2008.00482.x.

Warren, D. L., and S. N. Seifert. 2011. "Ecological Niche Modeling in Maxent: The Importance of Model Complexity and the Performance of Model Selection Criteria." *Ecological Applications* 21(2): 335–42. https://doi.org/10.1890/10-1171.1.

Williams, K. J., L. Belbin, M. P. Austin, J. L. Stein, and S. Ferrier. 2012. "Which Environmental Variables Should I Use in My Biodiversity Model?" *International Journal of Geographical Information Science* 26(11): 2009–47. https://doi.org/10.1080/13658816.2012.698015.

Wright, G. D., and J. K. Frey. 2015. "Habitat Selection by the Endangered New Mexico Meadow Jumping Mouse on an Irrigated Floodplain." *Journal of Fish and Wildlife Management* 6(1): 112–29. https://doi.org/10.3996/062014-JFWM-044.

Zarzo-Arias, A., M. del Mar, S. P. Delgado, I. A. Jordana, G. Bombieri, E. González-Bernardo, A. Ordiz, C. Bettega, R. García-González, and V. Penteriani. 2020. "Seasonality, Local Resources and Environmental Factors Influence Patterns of Brown Bear Damages: Implications for Management." *Journal of Zoology* 313(1): 1–17. https://doi.org/10.1111/jzo.12839.

Zarzo-Arias, A., V. Penteriani, M. Del Mar, P. P. Delgado, R. G.-G. Torre, M. C. Mateo-Sánchez, P. V. García, and F. Dalerum. 2019. "Identifying Potential Areas of Expansion for the Endangered Brown Bear (*Ursus arctos*) Population in the Cantabrian Mountains (NW Spain)." *PLoS One* 14(1): 1–15. https://doi.org/10.1371/journal.pone.0209972.

Zedrosser, A., O. G. Støen, S. Sæbø, and J. E. Swenson. 2007. "Should I Stay or Should I Go? Natal Dispersal in the Brown Bear." *Animal Behaviour* 74(3): 369–76. https://doi.org/10.1016/j.anbehav.2006.09.015.

Zhang, Y., J. Tang, G. Ren, K. Zhao, and X. Wang. 2021. "Global Potential Distribution Prediction of *Xanthium italicum* Based

on Maxent Model." *Scientific Reports* 11(1): 1–10. https://doi.org/10.1038/s41598-021-96041-z.

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(1): 3–14. https://doi.org/10.1111/j.2041-210x.2009.00001.x.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Zarzo-Arias, Alejandra, Vincenzo Penteriani, Lukáš Gábor, Petra Šímová, Florencia Grattarola, and Vítězslav Moudrý. 2022. "Importance of Data Selection and Filtering in Species Distribution Models: A Case Study on the Cantabrian Brown Bear." *Ecosphere* 13(12): e4284. https://doi.org/10.1002/ecs2.4284