Environmental Management
https://doi.org/10.1007/s00267-024-01941-1

Applying XGBoost and SHAP to Open Source Data to Identify Key Drivers and Predict Likelihood of Wolf Pair Presence

Jeanine Schoonemann[1] · Jurriaan Nagelkerke[1] · Terri G. Seuntjens[1] · Nynke Osinga[2] · Diederik van Liere[2]

Received: 5 September 2023 / Accepted: 20 January 2024

Abstract
Wolves have returned to Germany since 2000. Numbers have grown to 209 territorial pairs in 2021. XGBoost machine learning, combined with SHAP analysis is applied to predict German wolf pair presence in 2022 for 10 × 10 km grid cells. Model input consisted of 38 variables from open sources, covering the period 2000 to 2021. The XGBoost model predicted well, with 0.91 as the AUC. SHAP analysis ranked the variables: distance to the closest neighboring wolf pair was the main driver for a grid cell to become occupied by a wolf pair. The clustering tendency of related wolves seems to be an important explanatory factor here. Second was the percentage of wooded area. The next eight variables related to wolf presence in the preceding year, except at fifth, eighth and tenth position in the total order: human density (square root) in the grid, percentage arable land and road density respectively. Other variables including the occurrence of wild prey were the weakest predictors. The SHAP analysis also provided crucial added value in identifying a variable that had threshold values where its contribution to the prediction changed from positive to negative or vice versa. For instance, low density of people increased the probability of wolf pair presence, whereas a high density decreased this probability. Cumulative lift techniques showed that the model performed almost four times better than random prediction. The combination of XGBoost, SHAP and cumulative lift techniques is new in wolf management and conservation, allowing for the focusing of educational and financial resources.

Keywords XGBoost · SHAP · Ecological model · Machine learning model · Geospatial analysis · Wolf

Introduction

The wolf (*Canis lupus*) is rapidly recolonizing Europe (Chapron et al. 2014; Kuijper et al. 2016), with the wolf population in Germany having a key role in this development (Jarausch et al. 2021). In 2000, the first successful reproducing wolf pair was observed at the Muskauer Heide in Saxony (Mideast Germany) (DBBW 2023). The presence of wolf pairs in German territories has been collected and documented by the German Dokumentations- und Beratungsstelle des Bundes zum Thema Wolf (DBBW 2023). Their data show a steady increase in the number of wolf pair territories, with 209 wolf territories in 2021. After the recolonization in Germany, reproducing wolves settled in Denmark in 2017 (Naturhistorisk museum Aarhus 2023), The Netherlands in 2018 (Jansman et al. 2021) and Belgium in 2019 (Gouwy et al. 2019). A wolf pair lives and reproduces in a territory which provides food and rest. The young wolves may leave the parental territory to settle elsewhere (Mech and Boitani 2003). Wolves may inhabit areas close to human settlements and roam in these areas. In the countries mentioned above, settled wolves are seen during the day and within built-up areas. In addition, wolves may kill livestock, though 18% of identified wolves in the Netherlands did not (van Liere et al. 2021). The proximity of a large predator to humans and livestock leads to social unrest and intolerance towards the protection of wolves. It is therefore important to understand the wolves' preferences for settlement and to predict where that might happen in new areas. Machine learning models are useful to this end, but there is often a trade-off between the predictive power and the transparency of the analyses. Therefore, this study proposes a novel approach which resolves this trade-off, namely by combining a model with high predictive power, such as XGBoost (eXtreme Gradient Boosting) with a SHAP (Shapley Additive Explanations exPlanations)

✉ Jeanine Schoonemann
j.schoonemann@cmotions.nl

[1] Cmotions, Kosterijland 40, 3981 AJ Bunnik, Nederland

[2] Institute for Coexistence with Wildlife, Heuvelweg 7, 7218 BD Almen, Nederland

Published online: 19 February 2024

analysis that provides the possibility of interpreting the importance and influence of variables.

A growing body of literature describes the use of machine learning in various ecological subfields due to their utility in capitalizing on big data and their potential for understanding the processes underlying ecological patterns (Tuia et al. 2022). Artificial Neural Network (ANN) modeling has for instance been applied to wolf distribution patterns in Portugal (Bessa-Gomes and Petrucci-Fonseca 2003). However, ANN models may have difficulty rendering absence data (areas without wolves recorded) and only work on small datasets when complying with specific requirements and data preprocessing (Pasini 2015). The maximum entropy-based machine learning model MaxEnt has a solution to these problems; it is capable of only working with recorded presences. This MaxEnt machine learning model has been applied to the distribution of wolves in Central Italy (Bassi et al. 2015), Germany (Kramer-Schadt et al. 2020) and in the USA in an ensemble of various models (Gantchoff et al. 2022; van den Bosch et al. 2022). One of this model's drawbacks lies in its limitation on the number of variables when working with data characterized by a low record (presence) count. A high ratio of input variables to records could pose challenges for the model, potentially causing it to emphasize irrelevant specifics while struggling to discern meaningful patterns (Halvorsen 2013). Furthermore, ecologists need to identify environmental or temporal variables that influence ecological patterns linked to distribution data, which implies that the interpretability of the model is crucial. To be able to interpret the results of the MaxEnt model, extensive preprocessing of the data is required. It expects users to minimize correlation among predictors and to identify variable-target-relationship shapes prior to model building. This is caused by the fact that the complex variables created by MaxEnt are often already highly correlated (Bassi et al. 2015). In this study, XGBoost is suggested as an alternative, as it does not require extensive preprocessing of variables and has few limitations regarding the number of variables added to the model, as well as their correlations. This study in ecological modeling takes advantage of XGBoost's quality and specifically uses a large number of variables based on open sources. Only open data sources and open source packages were used in our code to build the model; this code is also publicly available (https://dev.azure.com/Cmotions/Projects/_git/predict-wolf-presence). This approach has the added advantage that the methods are fully transparent and accessible, and the results can be easily reproduced or extended.

XGBoost is a powerful and widely used supervised machine learning algorithm that belongs to the ensemble learning category. In ensemble learning the final model is composed of multiple separate models. Furthermore, XGBoost is based on the gradient boosting framework and aims to create a robust and accurate predictive model. XGBoost employs a boosting technique that sequentially combines decision trees in an additive manner. It optimizes a loss function by iteratively minimizing the residuals of the previous model, meaning each new decision tree focuses on improving on the weakness of the previous decision tree, effectively improving the overall model's performance. XGBoost incorporates regularization techniques to mitigate overfitting and enhance generalization and employs advanced features like parallel computing and tree pruning to improve efficiency. XGBoost is highly regarded for its exceptional predictive accuracy and its ability to handle diverse types of data, including spatial and temporal information (Chen and Guestrin 2016). Thus, XGBoost is used in this study, but still, like many other advanced machine learning models, XGBoost is a black box. Therefore, XGBoost is combined with a powerful framework for interpreting predictive models: SHAP.

SHAP is a method used to explain predictions of various machine learning models in a uniform way. It assigns importance values to each input variable. The framework is based on game theory concepts, specifically the idea of Shapley values, which allocate contributions to each player in a cooperative game. Similarly, when used in ecology, it provides variable importance measures that allow researchers to interpret the contributions of individual variables towards the model predictions (Lundberg and Lee 2017).

To evaluate the predictive power and usability of the model created by XGBoost, the Receiver Operator Curve (ROC) and its Area Under the Curve (AUC) are used. The ROC plots the True Positive Rate versus the False Positive Rate, whereas the AUC metric summarizes the model performance in one statistic. ROC and AUC are used to compare overall predictive performance between models and with expectations without a predictive model (Kuhn and Johnson 2013). Since the ROC and AUC only provide an overall performance score, they don't reveal how well the model performs on specific segments of the grid cells. Therefore, the cumulative gains and cumulative lift plots were added. These are techniques used in branches like marketing and finance to improve the evaluation of the predicted values by the model (Nagelkerke 2022). Cumulative lift and gains allow for a better understanding of the full distribution of model probabilities in a population, identifying segments in the population where the model performs exceptionally well. These techniques indicate to what extent the actual values (in our case, the presence of wolf pairs) are within the population segments with the highest predicted probabilities (cumulative gains) and how much better the model performs for those segments as compared to not using any model (cumulative

lift). These insights can be used to turn into actionable insights, for instance, in marketing, and can help to identify segments of the customer base where a marketing campaign will be most effective, improving resource allocation. In our study, this would translate to selecting a specific subset of grid cells with high probabilities of predicted wolf pair presence. Such selection can be motivated, for instance for future management and resource allocation in specific areas.

This study aims to demonstrate the significance of the combination of XGBoost and SHAP modeling, as well as the importance of model evaluation techniques, to improve the prediction of the presence of wolf pairs in Germany, which is essential for wolf management and protection.

Methods

Multivariate modeling and analysis of variables possibly contributing to the prediction of wolf pair settlement in Germany was performed using XGBoost (XGBoost package 1.7.5 2023, https://pypi.org/project/xgboost/) combined with SHAP calculations (SHAP package 0.41.0; 2023, https://pypi.org/project/shap/; Lundberg et al. 2020). In addition, common exploratory data analysis regardless of correlations with other variables was performed by calculating box plots for each variable.

Model

The XGBoost model is used from the Python packages xgboost and scikit-learn (Pedregosa et al. 2011; Scikit-learn package 1.2.2 2023, https://pypi.org/project/scikit-learn/). The hyperparameters are Bayesian optimized with the Python package hyperopt (Bergstra et al. 2015; Hyperopt package 0.2.7 2023, https://pypi.org/project/hyperopt/). Optimal values of the hyperparameter optimization as used in the XGBoost model are given in Table 1. The Python code of this study is online available (https://dev.azure.com/Cmotions/Projects/_git/predict-wolf-presence).

To ensure a systematic evaluation of the model's performance while preserving the integrity and representativeness of the data, the produced dataset is split into three parts: training, testing, and validation. The training and testing dataset cover all data from the grid cells for the period 2000 until 2019, where this data is split randomly such that 75% of the data is part of the training dataset and 25% is part of the testing dataset. The validation data set covers all data for the period 2020 until 2021. Stratified K-fold cross-validation with ($K = 10$) was used. For validation the ROC curve with the related AUC metric and the Cumulative Gains and Lift Curves from the Python package modelplotpy (Modelplotpy package 1.0.0 2023, https://modelplotpy.readthedocs.io/en/latest/) were used.

Analysis of the relative significance and effect of the applied variables on the model's prediction was done with SHAP methodology from the eponymous SHAP Python package. With SHAP, we can identify the corrected relationship between input and target variable, showing that the relationship can be both positive and negative depending on the input variable's value.

Grid cells

A map of Germany was loaded in QGIS, overlaid with a new layer with 10×10 km grid cells. This new layer was retrieved from the EUROSTAT data (open source data EUROSTAT 2020, https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/countries, accessed 2023). The combination was exported as a new single geospatial (shape) with Germany divided into grid cells ($n = 3867$). The exact grid cell area was calculated for those grid cells ($n = 564$) that were located adjacent to the foreign border and the German territorial seas.

Presence of wolves

For the presence of wolf pairs in German territories data was used that was collected by the DBBW since 2000 (open

Table 1 Optimal values of hyperparameters with the applied search space for XGBoost modeling in this study

Hyperparameter	Search space	Optimal value
learning_rate	log-spaced array: 1000 values ranging from 0.005 to 0.5	0.0494
max_depth	sequence of integers ranging from 5 to 31 (inclusive); step size 1	17
min_child_weight	uniform distr. of discrete values between 1 and 10 (inclusive)	5
gamma	[0.5, 1, 1.5, 2, 5]	1
subsample	uniform distr. of values between 0.1 and 1 (inclusive); step size 0.01	0.64
n_estimators	discrete values from 20 to 200 (inclusive); step size 5	34
colsample_bytree	uniform distr. of values between 0.1 and 1 (inclusive); step size 0.01	0.72
reg_alpha	[1e–5, 1e–2, 0.1, 1, 10, 100]	0
reg_lambda	[1e–5, 1e–2, 0.1, 1, 10, 100]	4

source data DBBW, https://www.dbb-wolf.de/home, accessed 2023). Recordings of DBBW also indicate whether there is reproduction, including data on the number of pups and yearlings, or whether there is a settlement of a single adult wolf or a non-reproducing pair. Wolf territories are indicated for each year by circles with a diameter of 16 km, resulting in an area of 200 km^2; the approximate size of a German wolf territory (Kramer-Schadt et al. 2020). The center of the circle is the center of a minimum convex polygon of repeated confirmed wolf presence in an area (Reinhardt and Kluth 2015; Reinhardt et al. 2017) and is considered the territorial center. DBBW only provides wolf territory locations as a geographical image. Therefore, recognizable map variables were determined that aligned with North-South and West-East lines through the circle center and were copied to Google Earth, which gave the latitude and longitude values at the cross of the lines. The estimation of the territory center was precise at 1 km. Subsequently, these estimations of territory location were associated with the grid cells. One grid cell might have been associated with multiple territories.

The target variable in this study is the first registered presence of a wolf pair in a grid cell, meaning that no other wolf pairs have been present in this specific grid cell before. Its value is either 1 (presence) or 0 (no presence). The presence of other wolves was analyzed in a spatial and temporal dimension. Spatially, distances were calculated between the grid cell center and the closest neighboring wolf pair in the same year. Temporally, a summation was performed of the number of adult wolves present in the year preceding the first presence in a grid cell within a 25 km radius from its center, which is the median dispersion distance of female wolves (Jarausch et al. 2021). The same was done for a 50, 75 and 100 km radius and was subsequently repeated for the number of pups, yearlings and the total number of wolves respectively.

Habitat variables

We aimed to maximize the number of possibly relevant variables included, to prevent missing a factor that may be relevant for wolves. The selected variables were based on previous studies in Germany and other European countries (Blanco et al. 1992; Massolo and Meriggi 1998; Jędrzejewski et al. 2000; Glenz et al. 2001; Eggermann et al. 2011; Ordiz et al. 2020).

The percent of a grid cell was calculated as being covered by arable land, artificial construction and sealed areas, bare surface, grassland, inland water, permanent crops, scrubs and wooded area, using the Land Use and Coverage Area frame Survey (open source data LUCAS from EUROSTAT, https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/countries, accessed 2023). LUCAS data points are associated with the intersections of a 2 km grid. These points were mapped onto the grid cells in this study and the percentage of total points within each grid cell was determined as an estimate of the percentage coverage. In addition, polygons defining Natura 2000 Areas (open source data European Environment Agency 2021, accessed 2023) were overlaid onto the grid cells and the percentage of a grid cells' area covered by Natura 2000 areas was calculated.

The human population density in each grid cell was retrieved from the Humanitarian Data Exchange (open source data HDX 2019, https://data.humdata.org/dataset/germany-high-resolution-population-density-maps-demographic-estimates, accessed 2023) The retrieved data represented the densities per km^2 for the year 2019. The assumption was that these were indicative of (differences between) grid cells across this study's whole period. All the data points that could be mapped to the corresponding grid cell were aggregated. Their values were summed, and the sum's square root was taken to estimate human population density in a grid cell.

The presence of German railways and roads was downloaded from DIVA-GIS (open source data DIVA-GIS 2011, http://www.diva-gis.org/gdata, accessed 2023). This data has not been updated since they were added in 2011. The assumption was that no significant change in road and railways presence had occurred since 2011 for most grid cells. In QGIS, the grid cell polygons were merged with the roads and railways(multi)line strings. Roads and railways were cut at the grid cell borders and the total length of (multi)line strings (in km) was calculated within each grid cell. The resulting export was a (shape)file suitable for Python, where we calculated the density of railways and roads per grid cell in km per km^2.

Estimations of the presence of wildlife species in a grid cell were based on observed occurrences in Germany as registered in the Global Biodiversity Information Facility (GBIF 2023, https://www.gbif.org/occurrence/download/0224656-230224095556074956074, accessed 2023). However, registrations were absent for some species for several years and lacked consistency for the wildlife species in this study until 2021. Therefore, data were selected and cumulated, covering observed occurrences between January 2021 and April 2023. There was a variation in the precision of these latitude and longitude registrations, and it was also considered that spotted animals are mobile. Uncertainties larger than 5 km involved between 0.5 and 17.5% of the data. Thus, 82.5% of the locations range within 5 km. Therefore, a circle was adopted with a radius of 5 km from the registered location. These circles were superimposed onto the grid cells. An animal's presence was then defined as the percentage of overlap for a given grid cell. Due to the mentioned limitations of this open source data, we were only able to approximate the presence of wildlife in a grid cell.

Wildlife prey of German wolves is mainly roe deer (*Capreolus capreolus*), but also other species are predated such as red deer (*Cervus elaphus*), wild boar (*Sus scrofa*), mouflon (*Ovis musimon*), and European hare (*Lepus europaeus*) (Ansorge et al. 2006). Polish studies mention fallow deer (*Dama dama*), Alpine ibex (*Capra ibex*) and beavers (*Castor fiber*) (Nowak et al. 2011; Reinhardt et al. 2021; Witek et al. 2023). Densities of roe deer and red deer were found to be the main drivers for wolf settlement in the Western Alp region of Switzerland (Roder et al. 2020). All European species with the same genus as mentioned in these sources were selected in the GBIF database. However, data on beavers (*Castor fiber*) was not available in the GBIF database. The raven, *Corvus corax*, was also selected as wolf and raven play together and facilitate each other's prey or carcass finding (Stahler et al. 2002; Erdas 2020). Densities of domestic species were not included, as they could not be retrieved from an open source database in Germany.

Results

Model

The training, testing, and validation datasets were comprised of 38 input variables (appendix A in the supplementary materials) and the target variable. The training and testing dataset covered the data from 2000 to 2019 and included data for 180 grid cells with wolf pair occupation. The validation dataset covered the data in the last two years, 2020 and 2021. In this period, there were 43 grid cells with first time wolf pair occupation. Therefore, the total number of grid cells with past or current presence(s) reached a cumulative number of 223 grid cells in 2021. The AUC of the optimized XGBoost model is 0.91 for the test data and 0.81 for the validation data (Fig. 8 in the supplementary materials).

Grid cells and the presence of wolves

A total of 1180 counts of wolf pair presence applied for the period 2000–2021. In many cases the presence was found in the same grid cells, resulting in 223 unique grid cells with wolf pair settlements. The prediction of the presence of wolf pairs by the XGBoost model is provided in Fig. 1. It shows for each grid cell the probability of wolf pair settlement in 2022, the year following the research period. The 223 grid cells where wolf pairs were already present before were not included in the prediction but are presented as gray dots instead.

The predicted presence in 2022 does not expand strongly towards the far western region (near Netherlands and Belgium) or northern Germany (near Denmark) but remains dominant in mid-northern and northeastern Germany: Sachsen, Sachsen-Anhalt, Brandenburg, eastern Niedersachsen and Mecklenburg-Vorpommern (Fig. 1). The prediction also shows that the isolated groups of wolf pair presence in the southern half of Germany expand.

During the years 2020 and 2021, a selection of 10% of the grid cells exhibiting the highest in probability of wolf pair occupancy, represented approximately 35% of all grid cells occupied by wolf pairs (Fig. 2). Conversely, when no model would have been used, it could be expected that 10% of the cells would show occupancy of 10% of all wolf pairs. This results in a cumulative lift (Fig. 9 in the supplementary materials) of about 375% for the 10% grid cells with the highest probability of wolf pair presence. Thus, for these grid cells, the model performs almost four times better than in the case of random selection of areas.

Habitat variables and SHAP values

Average SHAP values were established for all variables in the XGBoost model (Fig. 3), which shows the relative importance of these variables for the prediction. The distance to the closest neighboring pair of wolves and the percentage of wooded area are most important in predicting the first presence of a pair of wolves in a grid cell. The next eight all relate to presence of wolves in the preceding year, except three on fifth, eighth and tenth position in the total order: the human population number (square root) in the grid, percent of arable land and road density respectively. Then there are several variables, including those related to wildlife species, that result in SHAP values that are among the lowest and therefore the weakest predictors.

The influence of the distance to the closest neighboring pair of wolves can be both positive and negative, depending on the value for this variable. The influence is positive for small distances approximately between 10 and 40 km but negative for larger ones (Fig. 4). SHAP values for distances larger than 100 km can be found in Fig. 10 in the supplementary materials.

Woodland cover in a grid cell can also have both a positive and a negative impact on wolf pair presence. The SHAP values are positive for coverage higher than approximately 40%. Such percentages of cover therefore positively impact the probability that wolf pairs will occupy that area (Fig. 5). Lower cover percentages correspond to negative SHAP values and reduce this probability. The percentage of arable land cover mirrors this result. Here, SHAP values are positive with arable land cover lower than 40% (Fig. 11 in supplementary materials).

Examples of the impact of presence of other wolves on SHAP values, that is of adult wolves within a 25 km radius (Fig. 12 in the supplementary materials) and pups within a 50 km radius (Fig. 13 in the supplementary materials) preceding the year of first presence of a wolf pair in a grid cell,

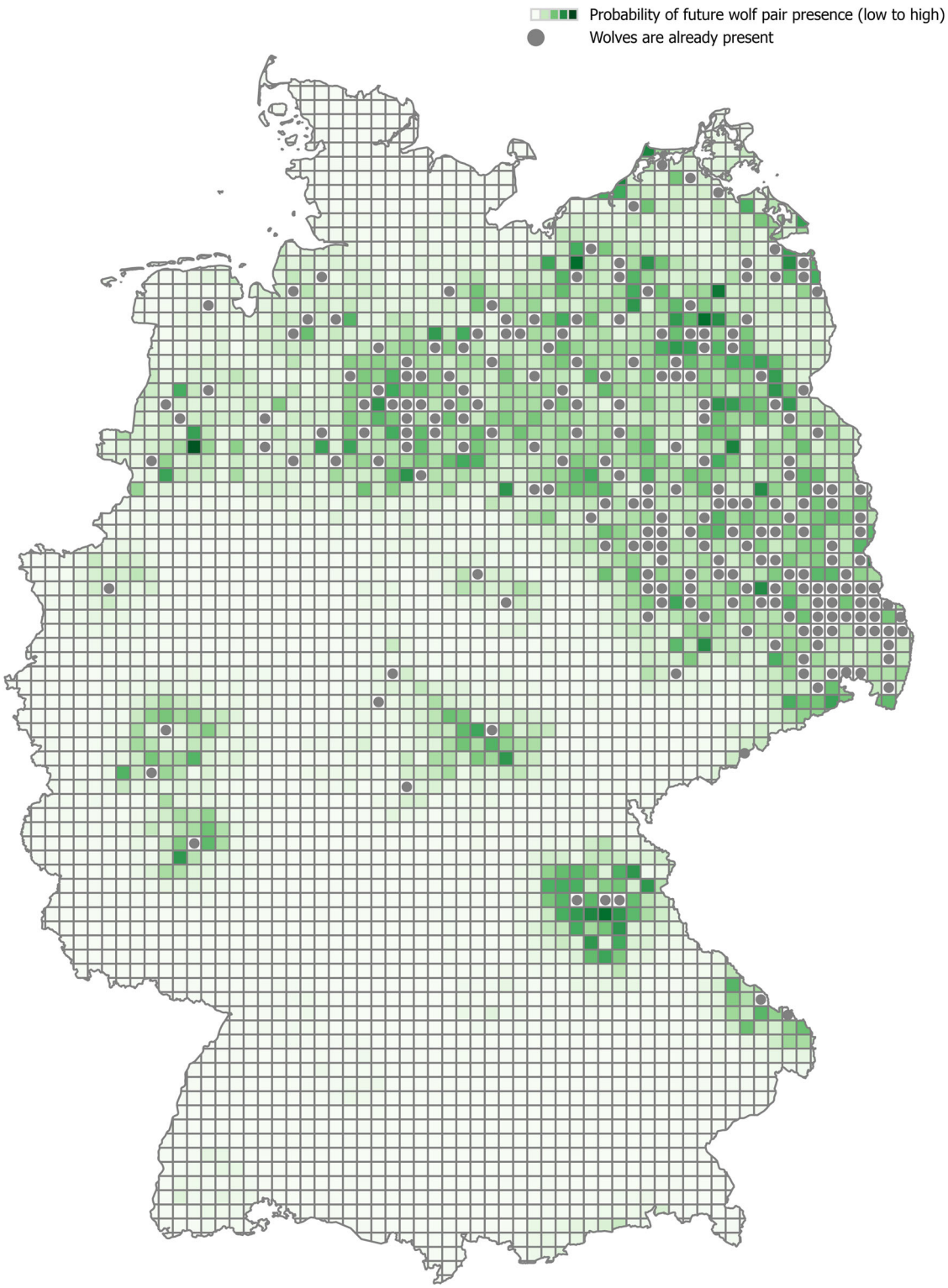

Fig. 1 Probability of wolf settlements in grid cells covering Germany for the year 2022 (greener corresponds to a higher probability) for grid cells with no previous wolf pair presence. Grid cells with a known previous presence of wolf pairs are also presented (gray dot)

show positive influences on the probability of subsequent occupation of this grid cell, though a negative one for less than 10 pups within a 50 km radius.

Human presence has both a positive and a negative effect on chances of wolf pair presence (Fig. 6). If the square root of human population density in a grid cell is lower than 80

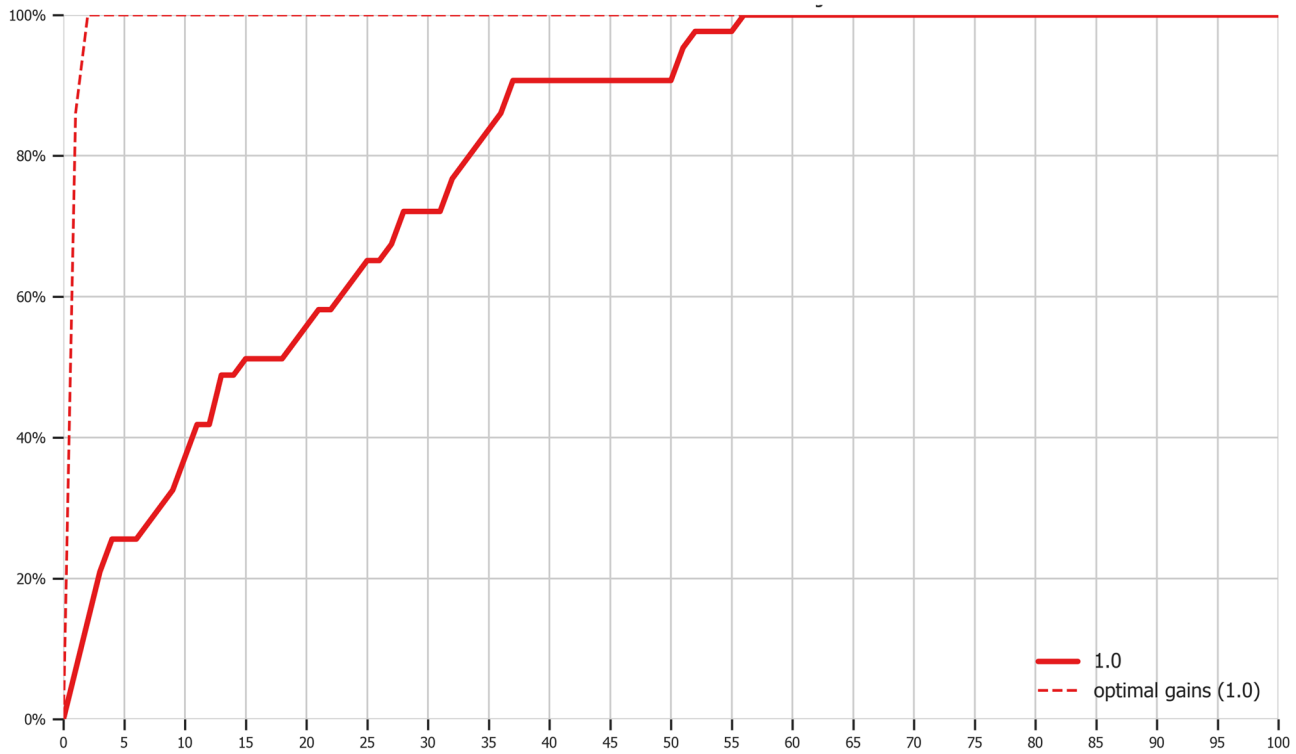

Fig. 2 The percentage of actual presence of a wolf pair as cumulative gains of the XGBoost model in this study (solid line) and a perfect model (dashed line) as related to the percentage of grid cells ranked by the model's predicted probability of wolf pair presence from high to low

per grid cell, which can be translated to a human population density of 64 per km^2, the SHAP value is positive and the probability for a wolf pair to be present is enhanced. Higher human densities relate to negative SHAP values, implying reduced chances of wolf pair presence.

Regarding infrastructural variables, relatively weak SHAP values for road densities and one of the lowest values for railway densities are found. Road densities of 0.16 km per km^2, and higher have a negative impact on SHAP values and therefore on the probability of presence of a wolf pair (Fig. 7).

Observed occurrences of wildlife species differ strongly (Table 2), but roe deer, which wolves strongly prefer as prey, is spotted the most. Also, hares and ravens are frequently spotted.

Compared to other wildlife species, the occurrence of roe deer in a grid cell has on average the highest average SHAP value, though it still is a weak predictor. The SHAP values are positive and therefore contributing to the probability of presence of a wolf pair in a grid cell if the relative area in which a roe deer is seen is less than 5% (Fig. 14 in the supplementary materials). Conversely, the SHAP values are negative and reduce the probability of the presence of a wolf pair if the relative area in which a roe deer is seen is greater than 40%.

Habitat variables and box plots

For each input variable in the training dataset, across the years 2000–2019 boxplots were calculated for grid cells with and without a wolf pair present (target value 1 and 0 respectively) (Fig. 15 in the supplementary materials).

Comparison of boxplots, particularly medians and interquartile distance, indicates that for grid cells where wolf pairs are or have been present, the distance to another wolf pair is smaller compared to grid cells without a wolf pair. In the year preceding occupation of a grid cell, the number of wolves (total and per age category: pups, yearlings, adults) is generally higher than random at different distances from the grid cell. The percentage of cover with forest or Natura 2000 is greater in grid cells with a (former) wolf pair occupancy than in grid cells without. The reverse is true for the percentage of arable land cover and the (square root) of the human population. The presence of prey animals is either not different between cells with or without presence of wolf pairs or was lower in cells with wolf pair presence (roe deer and hare). When all prey animals are taken together, their presence is lower in case a wolf pair occupies a grid cell. The presence of raven or densities of roads or railways do not differ between cells with or without wolf pairs present.

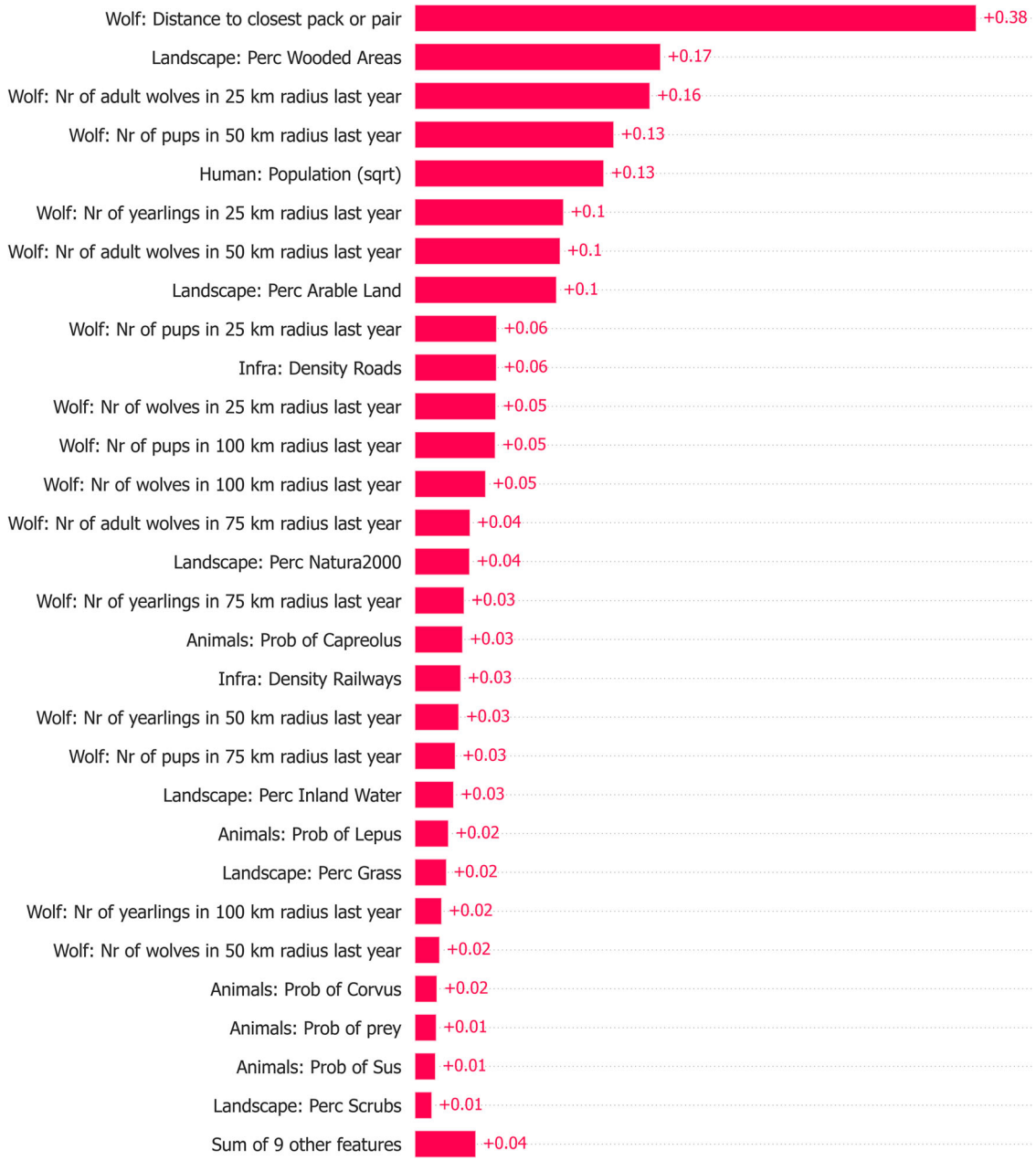

Fig. 3 Average SHAP values for all variables used in the XGBoost model relating to their relative significance for the model prediction

Discussion

Model

The XGBoost model used in this study included 38 variables (appendix A in the supplementary materials) based on open source data sources and performed well with an AUC of 0.91 on the test data. This is a high AUC value, especially when considering that we worked with a binary-dependent variable covering the settlement of wolf pairs in 223 (6%) out of 3867 grid cells. Moreover, this performance estimate is distinctly better than for instance the 0.76 of the MaxEnt model predicting German wolf distribution with cross-validation, as used by Kramer-Schadt et al. (2020). Thus, this study shows that a preference for XGBoost is justified, as it requires very little data preparation and has no real limitations regarding the number of input variables. Our analyses also show that the model properly generalizes (AUC value of 0.81 for the validation dataset) and is therefore capable of predicting occupation for areas in periods it is not trained on. The predicted occupation by wolf pairs in 2022 remains

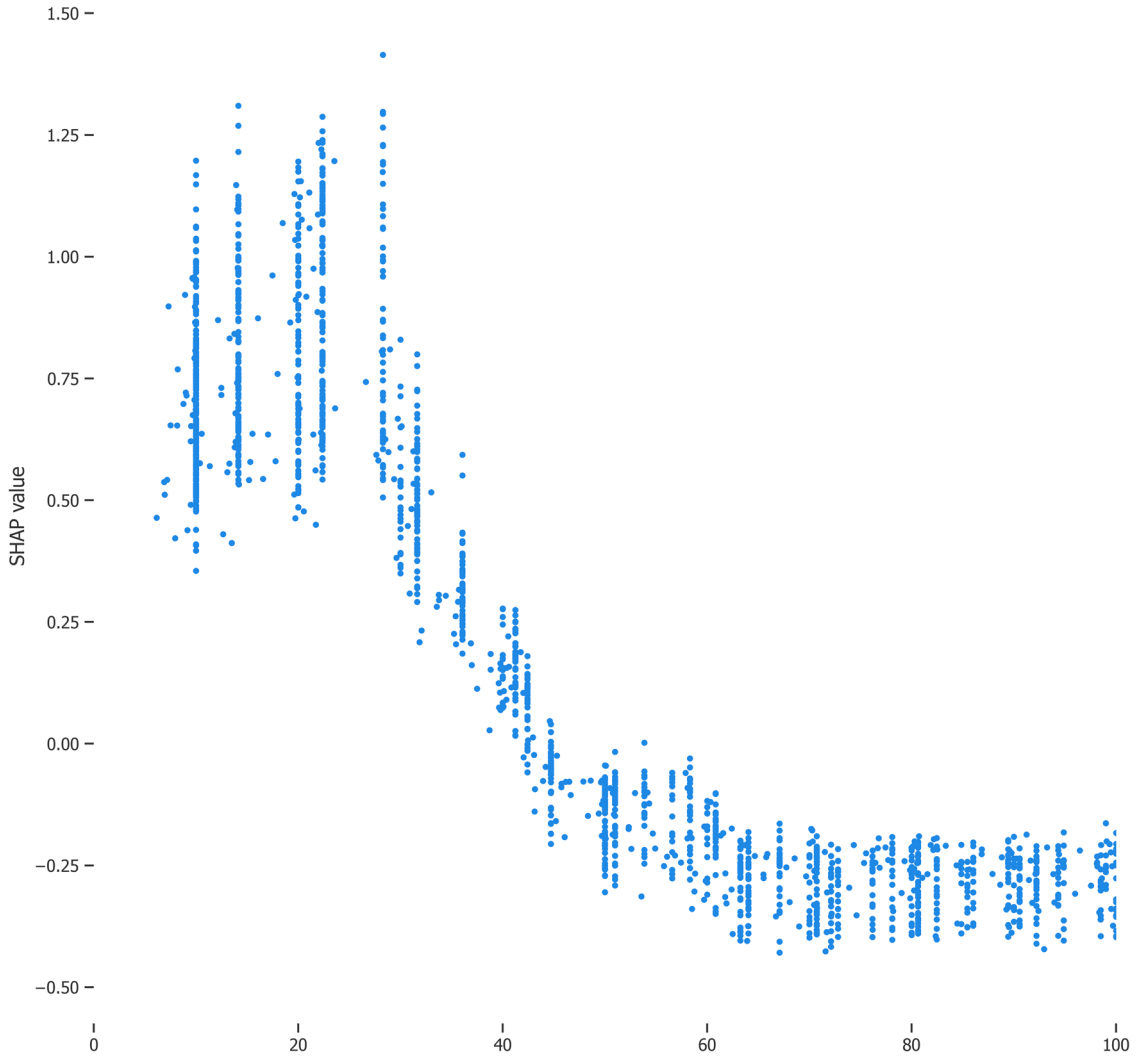

Fig. 4 The impact of the distance (in km) of other wolves to the settlement of a new wolf pair on SHAP values

dominant in mid-northern northeastern Germany (Sachsen, Sachsen-Anhalt, Brandenburg, eastern Niedersachsen and Mecklenburg-Vorpommern), while isolated groups in the southern half of Germany may expand. For these areas the cumulative gains and cumulative lift plots prove useful in evaluating the predictive quality of the model for the areas with the highest presence probabilities according to the model. The plots show that when the model is used to select the top 10% grid areas with the highest probability, these areas account for 35% of all expected presence of wolf pairs in the coming period and that for these areas, the prediction is almost four (3.75) times better than randomly selected grid areas. Therefore, wolf management, resources or research programs can be more focused in the mentioned states and areas in Germany. For instance, involvement of the darkest green grid cells in Fig. 1 may support management to locally educate people about coexistence with wolves, to support farmers to protect husbandry animals against wolf attacks, or to research development of habits, habitat use, and relations between wolf packs.

SHAP

The SHAP application has had an important added value, as it ranks the significance of variables for the black box model and illustrates how they contribute to the prediction. SHAP analysis also allows to identify threshold values to discriminate between positive and negative effects, This pattern could not as easily have been detected by a standard comparison, for instance of variable estimates between grid cells with or without wolf pairs or by linear modeling of variables. When using other techniques, such as regression-based approaches, much more effort is required to model the impact of variables. One needs to explicitly specify the shape of the relationship (linear, quadratic,..), interactions and isolate the effect of the variable from correlated

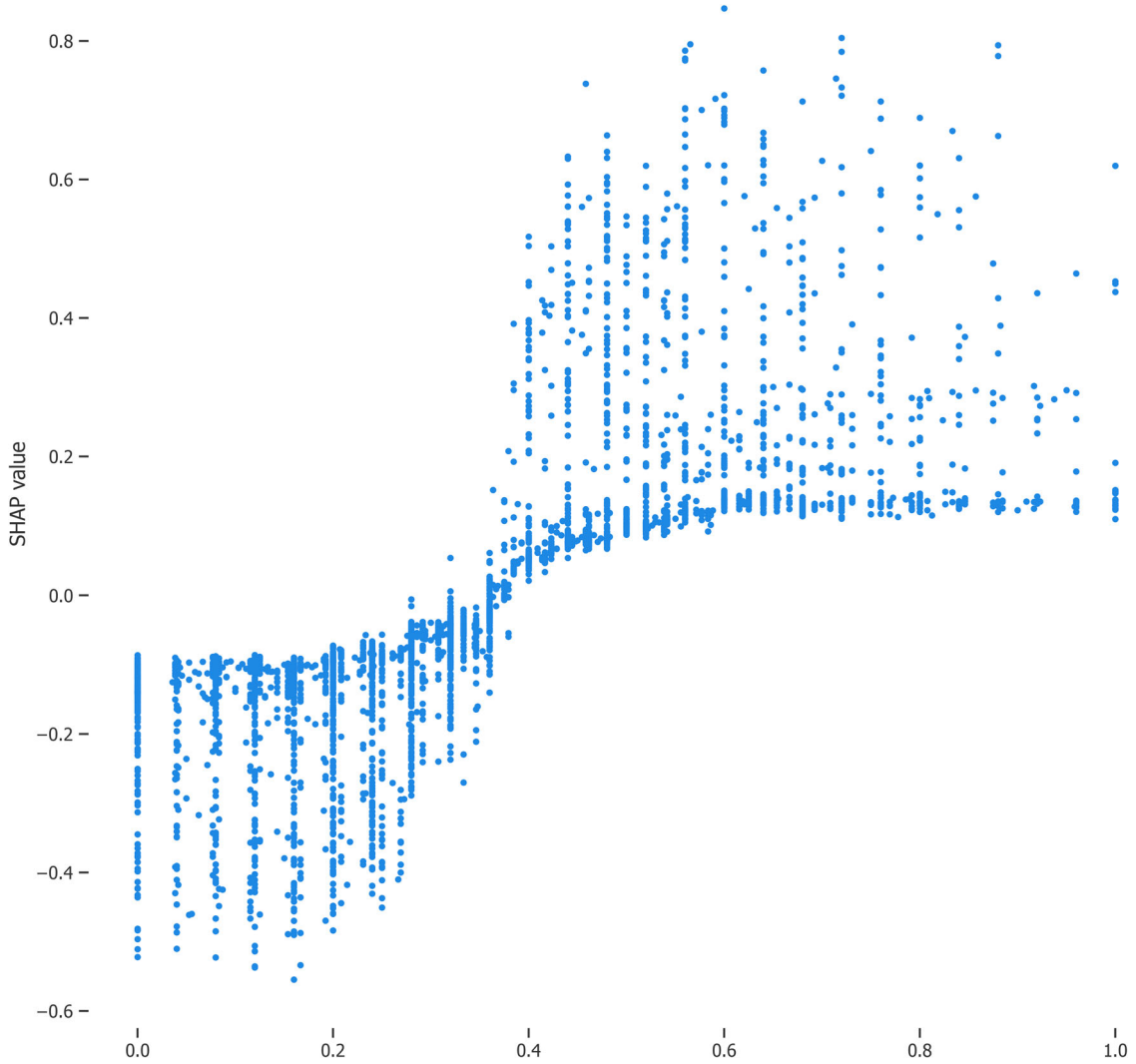

Fig. 5 The impact of the percentage of woodland cover in a grid cell on SHAP values

variables. In our approach – combining XGBoost with SHAP - nonlinear relationships between a variable and the target and correlations between variables are covered by design by XGBoost. SHAP enables interpretation of the relationship between variables and the target, controlling for the impact of other variables.

SHAP identified key drivers to predict the likelihood of wolf pair presence. The most important variable was the distance to the closest neighboring wolf pair. Although boxplots showed that short distances of cells with other wolves and the presence of wolves in the preceding year related to grid cells with a wolf pair, it is the SHAP analysis that values the magnitude of its significance for the prediction. A distance of about 10–40 km to a wolf pair already present contributes to a high probability that a wolf pair will be present in a previously empty grid cell. The significance of wolf presence was also confirmed by the findings of an increased chance of a grid cell occupation when wolves (adult, yearlings or pups) were present at 25 km a year before. For pups, a relatively strong SHAP value was also found for a distance of 50 km.

The presence of other wolves as a main variable supports the common notion that wolf packs tend to cluster over time (Mech and Boitani 2003). Patchiness of suitable habitat is often mentioned as one of the explanations for such clustering, especially the presence of forest. In our study, the forest cover percentage is indeed higher in grid cells with wolf pair establishment than in 'empty' grid cells and it has also the second highest average SHAP value. The significance of forest cover to the probability of wolf presence is consistent with other studies on the distribution of wolves in Europe (Massolo and Meriggi 1998; Jędrzejewski et al. 2000; Kramer-Schadt et al. 2020; Cimatti et al. 2021; Mayer et al. 2022; Marucco et al. 2023) and North America (Mladenoff et al. 1995, 2009; Oakleaf et al. 2006; Smith et al. 2016; Gantchoff et al. 2022). The SHAP analyses

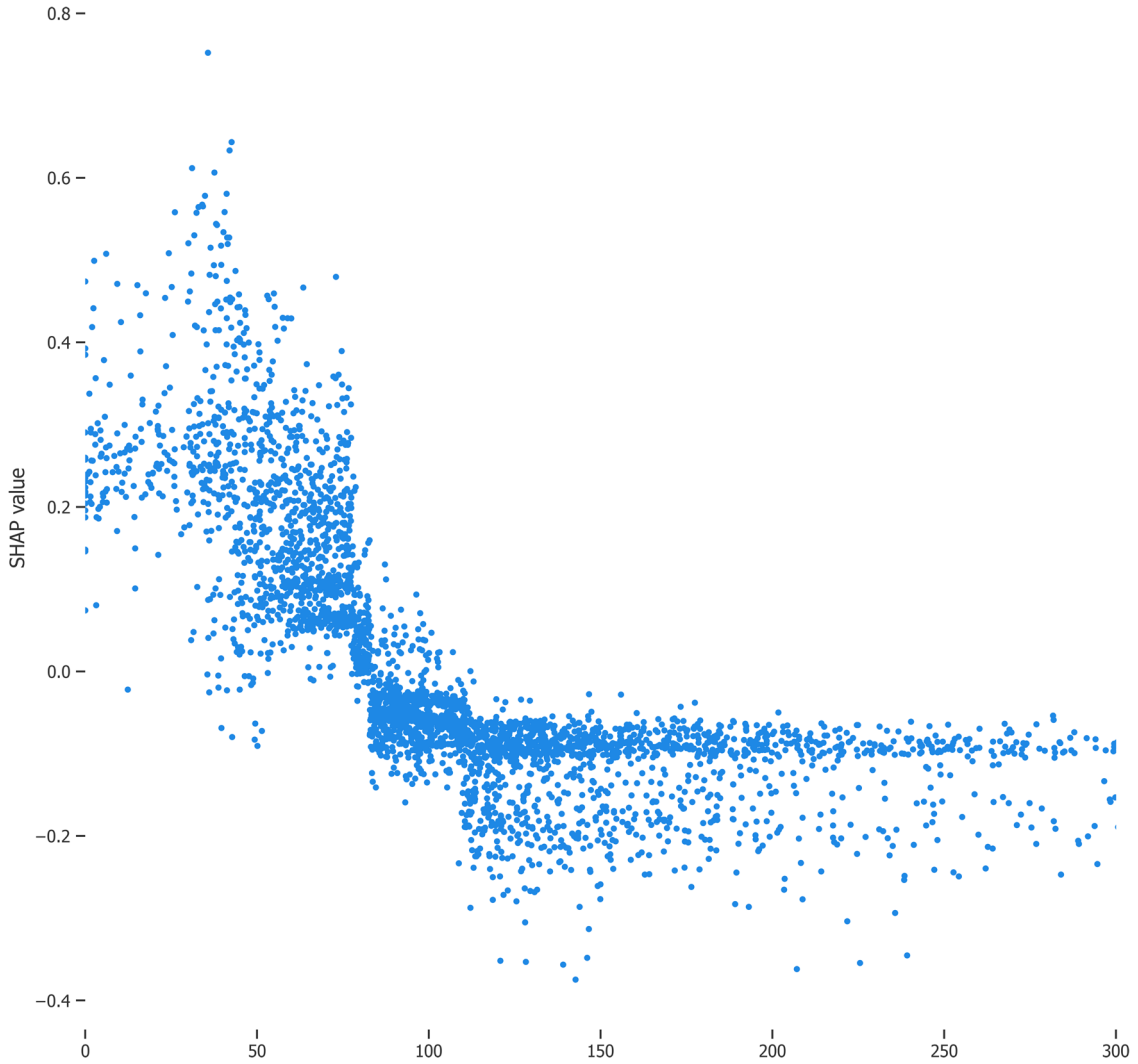

Fig. 6 The impact of the square root of human population numbers in a grid cell on SHAP values

conducted in the current study enabled quantification: probability of wolf pair presence starts to increase when more than 40% of the area is forest.

However, forest cover as an explanation for the clustering of wolves can be questioned, as other studies report that wolves inhabit a wide range of natural habitats (Mech and Boitani 2003; Oakleaf et al. 2006; Mladenoff et al. 2009; van den Bosch et al. 2022). The species is not specifically tied to specific habitats, but may preferably select forest areas and vice versa avoid arable land, areas with high road densities, or areas with high human densities, as this would relate to a reduced likelihood of human encounters (Mladenoff et al. 2009; Reinhardt and Kluth 2015; Reinhardt et al. 2019). It is also argued that wolves do not specifically prefer Natura 2000 protected areas and that forests do not play a key role in the recolonizing of Germany (Reinhardt and Kluth 2015; Reinhardt et al. 2019) and the northern Great Lakes region in the USA (Mladenoff et al. 2009). The estimation that forest cover is a less important habitat feature than human density is also assumed in other European and North American wolf studies (Smith et al. 2016; Cimatti et al. 2021; van den Bosch et al. 2022; Marucco et al. 2023). Mayer et al. (2022), however, could not identify human population density as a significant variable. Also in our study, human density scores below forest cover in average SHAP value and would therefore be less significant than wood cover in predicting wolf presence. Why this is, demonstrates the importance of adding a SHAP analysis: human density has both a positive and a negative impact on wolf pair settlement. At densities lower than 64 per km^2, the SHAP value is positive and the probability for a wolf pair to occupy a new area is even enhanced. Only a higher human population density of more than 64 per km^2 negatively affects the likelihood of an occupancy by a wolf pair in a new area. This is consistent with the results of the model of Van den Bosch et al. 2022,

Fig. 7 The impact of the density of roads in a grid cell in km per km^2 on SHAP values

Table 2 Number of occurrences of wildlife species that may relate to wolf presence observed in Germany between January 2021 and April 2023 as listed in the Global Biodiversity Information Facility

Species	Number of observed occurrences
Sus scrofa (wild boar)	1096
Ovis gmelini (mouflon)	97
Lepus europaeus (European hare)	6542
Lepus timidus (mountain hare)	2
Cervus elaphus (red deer)	520
Cervus nippon (sika deer)	19
Dama dama (fallow deer)	583
Capra ibex (Alpine ibex)	25
Capreolus capreolus (roe deer)	12,614
Corvus corax (common raven)	11,728

which estimated a sharp decline in wolf presence likelihood at human densities of 50 to 75 per km^2. The density of 64 people per km^2 is not extremely low in Germany and relates to rural areas with either much agriculture or nature areas. For comparison: the population density in Germany can be even lower: the districts (Landkreis) of Prignitz (Brandenburg), Altmarkkreis Salzwedel (Sachsen Anhalt) and Ostprignitz-Ruppin (Brandenburg) and Lüchow-Dannenberg (Niedersachsen) are listed as least populated in Germany with less than 40 inhabitants per km^2.

Concerning the clustering of wolves, even when assuming that patchy human population settlement (towns and villages) in Germany is most significant in predicting wolf settlement, it remains difficult to understand how this explains the significance of the specific distances to other wolves found in the current study (between 10 and 40 km

between locations where wolves are present, or of the presence of wolves in the preceding year within 25 km). Therefore, there is more to the prediction of wolf pair presence than human disturbance. Firstly, wolves may use their territory differently in successive seasons or years (Jędrzejewski et al. 2000) and therefore the center of the territory can move. This may (partly) explain the significance of the distances mentioned. In addition, cues by which wolves detect and recognize each other, such as from howling, scent marking in and at borders of the territory, may also explain such specific distances. Indeed, wolf kin tend to stay close: Caniglia et al. (2014) showed that within 17 km wolves are more closely related than on average across the population. Jarausch et al. (2021) found that in particular female dispersers or female offspring settled as breeding pack adjacent to the pack they left with a median dispersal distance of 25 km. The significance of distances to other wolves in the current SHAP analysis is in the range of the mentioned 17 and 25 km.

The SHAP analysis showed that road densities from 0.16 km per km^2 onwards would contribute negatively to the probability of wolf pair presence in an area. Road densities up to 0.7 to 1.1 appear to be possible for areas occupied by wolves (Mladenoff et al. 1995, 2009; Reinhardt et al. 2019; Gantchoff et al. 2022). However, road density had a relatively low contribution to the prediction. This result is also found by Reinhardt et al. (2019) and in other areas with expanding, recovering wolf populations like the north American western Great Lake areas (Gantchoff et al. 2022) or the northern Rocky Mountains (Oakleaf et al. 2006). The significance of road density is site specific (Oakleaf et al. 2006). Low road density in the American northern Great Lake areas used to be a key predictor for wolf settlement during first colonization (Mladenoff et al. 1995), but its significance declined after distribution expanded and wolves also settled in habitat that differed from the initial one (Mladenoff et al. 2009).

SHAP analyses also showed that the chance of observing prey animals relates to the weakest predictors for wolf pair presence, including roe deer. This result seems opposite to the findings that wolves live in a prey rich environment with wild ungulates as their primary prey (Jędrzejewski et al. 2000; Mech and Boitani 2003; Gazzola et al. 2005; Wagner et al. 2012; Kittle et al. 2017) and that wild prey can predict the presence of wolves (Oakleaf et al. 2006; Falcucci et al. 2013; Grilo et al. 2019; Roder et al. 2020). Nevertheless, when a prey species is abundant, there may be no difference in prey density between areas occupied by wolves compared to neighboring unoccupied areas. For example, Mladenoff et al. (1995) found both occupied and unoccupied areas had white-tailed deer (*Odocoileus virginanus*) densities of 8.4 to 8.6 per square km. Moreover, locations where prey can be caught easily can be more important than prey abundance itself (Zabihi-Seissan et al. 2022). In diet studies of wolves in or from Germany, roe deer is the main prey, but also red deer, wild boar, hares and rabbits (Wagner et al. 2012; Van der Veken et al. 2021; Jarausch et al. 2021). One explanation of the poor predictive power in the current study could be that prey is abundant across most of the grid cells. Germany is amongst the European countries with the highest roe deer densities, as estimations in 2005 were 8.4 animals per square km (Burbaitė and Csányi 2009). Also, the mere fact that wolves tend to cluster in Germany can relate to a general abundance of prey. A second explanation for the poor predictive power can be that prey can move from the center to the edge of a territory (Mech et al. 1980; Mech and Harper 2002). The avoidance behavior of the prey species is a known effect of wolf presence on the distribution of prey animals (Okarma 1995; Ripple et al. 2014; Kittle et al. 2017). A move of more than 5 km out of the core can result in prey presence outside the grid cell. As a result, cells not occupied by wolves but adjacent to the center grid cell can have increased prey animal densities compared to before wolf presence. This would be problematic in recognizing the actual importance of prey occurrence on the prediction of wolf pair occupancy in a specific grid cell. Indeed, the box plots showed that the chances of observing prey animals either do not differ between the cells with or without wolf pairs present or are lower in cells with a wolf pair. The latter applies in particular for roe deer and hare, which have been observed relatively frequently. It should, however, also be considered that the weak predictive power of the presence of prey animals can be a consequence of the quality of the open access data on prey species used in the model. There were inherent limitations, which are discussed below. Therefore, in this study, the predictive power of the occurrence of prey species should be interpreted carefully.

Open source data

Multiple sources of open data were used as input for the model. Open source data allows researchers from different disciplines to work on a broad range of research questions and build on each other's work (Roche et al. 2022). This study demonstrates that the application of a state-of-the-art model originating from the ´data science´ research field to open source ecology data, generates information from new angles, relevant to wolf conservation and management. Despite possible disadvantages of such data such as inconsistency and/or inaccuracy, as will be discussed below, the model created in this study is good and usable in practice: it has a high AUC on both test and validation data. Moreover, it showed good results when we validated it using data the model wasn't trained on.

The inconsistency of data applies to the presence of wildlife species. Due to inconsistent registration across years (years before 2021) in the Global Biodiversity Information Facility (GBIF 2023), the occurrence data were limited between January 2021 and April 2023. This is a period well after the last year used to train the model. To test the actual importance of prey densities in grid cells on the subsequent probability of wolf pair settlement, a follow-up model is suggested with annual data on prey densities. Moreover, since improved GBIF data have only been available and published since 2021, future research can incorporate GBIF data that is collected over a longer period of time (which is adequate for more extensive analyses) and deliver more accurate predictions of spatial and temporal relations between wolves and their prey.

The inaccuracy of data from open data sources has different origins. Sources can be outdated, for example regarding railway and road lengths in a grid. The effect of assuming negligible change across years is therefore not known until a comparison is made with annually updated sources. Sources may also be inaccurate and biased. For instance, the data sources used for prey and scavenger species may be biased by faults in human observations and varying degrees of observer coverage and intensity. Again, comparison and correlation studies with objective standardized monitoring (camera traps, systematic transect observations, trace density measurements etc.) are needed to resolve the extent of inaccuracy. In addition, the assessment of the center of a wolf pair territory was derived from a polygon of observations which led to inaccuracy depending on the number of observations. Its inaccuracy can be assumed to be larger than the subsequent estimation in the current study (with an accuracy of 1 km) of the center illustrated at the DBBW-site, but these errors have been systematic across all center assessments. Inaccuracies were also introduced by choosing a geospatial resolution of 10×10 km when sources were either on a more granular level but on a different projection or a less granular level and had to be attributed to multiple 10×10 km grid cells. It would be interesting to correct the mentioned inaccuracies and bias by enhanced measurements and to test to what extent the prediction improves.

The addition of not openly accessible data may also improve the prediction. For instance, the location of military areas relates to the stepwise distribution of wolves (Reinhardt et al. 2019). Moreover, the presence of domestic animals like sheep and goats is relevant, as these species are the main target of wolves that kill husbandry animals (Khorozyan and Heurich 2022). Also, the presence of beavers is of interest, as wolves may predate them too (Jędrzejewski et al. 2000; Reinhardt et al. 2019).

Further research

Further research into the application of SHAP to the XGBoost model is recommended, in particular regarding SHAP's ability to show variable significance even at the level of a specific area (or grid cell). Effects of variables on chances of wolf pair presence can significantly differ between areas. Figure 16 in the supplementary materials provides an example where a particular grid cell with a high probability (Fig. 16b in the supplementary materials) of wolf pair settlement provides a different order of the variables than another with a low probability (Fig. 16a in the supplementary materials). This local fine-tuning allows differentiation between variables that contribute to wolf settlement probability and enables management to focus on locally the most relevant ones.

Conclusions and perspectives

This study showed that XGBoost machine learning and SHAP analysis can be effectively applied to geospatial and temporal open source data. This approach is new and provides insight into input variable importance and quality regarding the prediction of wolf pair presence. XGBoost is amongst the best models according to literature, which is shown in our study as well, since it generated an AUC of 0.91. SHAP analysis explained that a short distance to another wolf pair is the most decisive variable predicting wolf pair presence. It also showed that variables, such as wooded area coverage and human population density, can contribute both positively and negatively to the prediction of the presence of a wolf pair. Road and prey densities, such as of roe deer, the wolves' primary prey, had poor contributions to the prediction of wolf pair presence. Out of a total of 3867 cells of 10×10 km covering Germany, 223 grid cells included presence of a wolf pair between 2000 and 2021. Management can be fine-tuned by selecting the 10% of grid cells with the highest predicted presence probability for the subsequent year 2022, which relates to a model performance almost four (3.75) times better than random. The prediction is therefore important to prepare for management instruments, such as education, and to locally prepare for coexistence with recolonizing wolves.

The strategy to combine XGBoost with SHAP analysis applied to predict wolf presence in Germany is promising as it contributes to the improvement of ecological modeling in general. The current model can also be directly used to predict wolf presence in countries that currently deal with and have conflicts with settling wolves, such as Denmark, The Netherlands, and Belgium. The preparation of data sets will be different between countries but will not alter the effectiveness of XGBoost and SHAP and their potential to contribute to management focused on promoting coexistence.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1007/s00267-024-01941-1.

Acknowledgements First and foremost, we would like to express our gratitude to the members of our research team, Simone Spierings, Kjeld Vissers, Philip Vermeij and Erik van Nistelrooij, who provided valuable input, insights, and assistance at every stage of the project. Their contributions were critical to the success of this research, and we are deeply grateful for their hard work and dedication. We also thank dr Markus Ritz for allowing us to use the DBBW (Dokumentations- und Beratungsstelle des Bundes zum Wolf) data, that are provided to the DBBW by the German federal states.

Author contributions JS: Methodology, Validation, Formal analysis, Writing—Original draft, Visualization, Project administration. JN: Methodology, Validation, Formal analysis, Writing—Review & editing. TS: Formal analysis, Writing—Review & editing. NO: Conceptualization, Writing—Review & editing. DvanL: Conceptualization, Writing—Original draft All authors reviewed the manuscript.

Compliance with ethical standards

Conflict of interest The authors declare no competing interests.

References

Ansorge H, Kluth G, Hahne S (2006) Feeding ecology of wolves Canis lupus returning to Germany. Acta Theriol (Warsz) 51:99–106. https://doi.org/10.1007/BF03192661

Bassi E, Willis SG, Passilongo D et al. (2015) Predicting the spatial distribution of wolf () breeding areas in a mountainous region of central Italy. PLoS One 10:e0124698. https://doi.org/10.1371/journal.pone.0124698

Bergstra J, Komer B, Eliasmith C et al. (2015) Hyperopt: a Python library for model selection and hyperparameter optimization. Comput Sci Discov 8:014008. https://doi.org/10.1088/1749-4699/8/1/014008

Bessa-Gomes C, Petrucci-Fonseca F (2003) Using artificial neural networks to assess wolf distribution patterns in Portugal. Anim Conserv 6:221–229. https://doi.org/10.1017/S1367943003003275

Blanco JC, Reig S, de la Cuesta L (1992) Distribution, status and conservation problems of the wolf Canis lupus in Spain. Biol Conserv 60:73–80. https://doi.org/10.1016/0006-3207(92)91157-N

Burbaitė L, Csányi S (2009) Roe deer population and harvest changes in Europe. Estonian J Ecol 58:169. https://doi.org/10.3176/eco.2009.3.02

Caniglia R, Fabbri E, Galaverni M et al. (2014) Noninvasive sampling and genetic variability, pack structure, and dynamics in an expanding wolf population. J Mammal 95:41–59. https://doi.org/10.1644/13-MAMM-A-039

Chapron G, Kaczensky P, Linnell JDC et al. (2014) Recovery of large carnivores in Europe's modern human-dominated landscapes. Science (1979) 346:1517–1519. https://doi.org/10.1126/science.1257553

Chen T, Guestrin C (2016) XGBoost. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, New York, NY, USA, pp 785–794

Cimatti M, Ranc N, Benítez-López A et al. (2021) Large carnivore expansion in Europe is associated with human population density and land cover changes. Divers Distrib 27:602–617. https://doi.org/10.1111/ddi.13219

DBBW (2023) DBBW, the Federal Documentation and Consultation Centre on Wolves. https://www.dbb-wolf.de/home. Accessed 15 May 2023

DIVA-GIS (2011) DIVA-GIS. http://www.diva-gis.org/gdata. Accessed 27 Jul 2023

Eggermann J, da Costa GF, Guerra AM et al. (2011) Presence of Iberian wolf (Canis lupus signatus) in relation to land cover, livestock and human influence in Portugal. Mamm Biol 76:217–221

Erdas C (2020) Wolves and Ravens: Defining a unique relationship. Osmosis Magazine

European Environment Agency (2021) Natura 2000 data - the European network of protected sites. In: https://www.eea.europa.eu/en/datahub/datahubitem-view/6fc8ad2d-195d-40f4-bdec-576e7d1268e4

EUROSTAT (2020) GISCO: GEOGRAPHICAL INFORMATION AND MAPS. In: https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/countries

Falcucci A, Maiorano L, Tempio G et al. (2013) Modeling the potential distribution for a range-expanding species: Wolf recolonization of the Alpine range. Biol Conserv 158:63–72. https://doi.org/10.1016/j.biocon.2012.08.029

Gantchoff MG, Beyer DE, Erb JD et al. (2022) Distribution model transferability for a wide-ranging species, the Gray Wolf. Sci Rep. 12:13556. https://doi.org/10.1038/s41598-022-16121-6

Gazzola A, Bertelli I, Avanzinelli E et al. (2005) Predation by wolves (Canis lupus) on wild and domestic ungulates of the western Alps, Italy. J Zool 266:205–213. https://doi.org/10.1017/S0952836905006801

GBIF (2023) GBIF, the Global Biodiversity Information Facility. In: https://www.gbif.org/occurrence/download/0224656-230224095556074

Glenz C, Massolo A, Kuonen D, Schlaepfer R (2001) A wolf habitat suitability prediction study in Valais (Switzerland). Landsc Urban Plan 55:55–65. https://doi.org/10.1016/S0169-2046(01)00119-0

Gouwy J, Van Den Berge K, Berlengee F, Mergeay J (2019) Wolvenspecial Oktober 2019. Roofdiernieuws Oktober:1–8

Grilo C, Lucas PM, Fernández-Gil A et al. (2019) Refuge as major habitat driver for wolf presence in human-modified landscapes. Anim Conserv 22:59–71. https://doi.org/10.1111/acv.12435

Halvorsen R (2013) A strict maximum likelihood explanation of MaxEnt, and some implications for distribution modelling. Sommerfeltia 36:1–132. https://doi.org/10.2478/v10208-011-0016-2

HDX (2019) Germany: High Resolution Population Density Maps + Demographic Estimates. In: https://data.humdata.org/dataset/germany-high-resolution-population-density-maps-demographic-estimates

Hyperopt (2023) hyperopt package 0.2.7. In: https://pypi.org/project/hyperopt/

Jansman HAH, Mergeay J, Van Der Grift EA, et al. (2021) De wolf terug in Nederland: een factfinding study. Wageningen

Jarausch A, Harms V, Kluth G et al. (2021) How the west was won: genetic reconstruction of rapid wolf recolonization into Germany's anthropogenic landscapes. Heredity (Edinb) 127:92–106. https://doi.org/10.1038/s41437-021-00429-6

Jędrzejewski WŁ, Jędrzejewska B, Okarma H et al. (2000) Prey selection and predation by wolves in Bialowieza primeval forest, Poland. J Mammal 81:197–212. https://doi.org/10.1644/1545-1542(2000)081<0197:PSAPBW>2.0.CO;2

Khorozyan I, Heurich M (2022) Large-scale sheep losses to wolves (Canis lupus) in Germany are related to the expansion of the wolf population but not to increasing wolf numbers. Front Ecol Evol 10: https://doi.org/10.3389/fevo.2022.778917

Kittle AM, Anderson M, Avgar T, et al. (2017) Landscape-level wolf space use is correlated with prey abundance, ease of mobility, and the distribution of prey habitat. Ecosphere 8: https://doi.org/10.1002/ecs2.1783

Kramer-Schadt S, Wenzler M, Gras P, Knauer F (2020) Habitatmodellierung und Abschätzung der potenziellen Anzahl von Wolfsterritorien in Deutschland. Deutschland/Bundesamt für Naturschutz
Kuhn M, Johnson K (2013) Applied Predictive Modeling. Springer New York, New York, NY
Kuijper DPJ, Sahlén E, Elmhagen B, et al. (2016) Paws without claws? Ecological effects of large carnivores in anthropogenic landscapes. Proceedings of the Royal Society B: Biological Sciences 283: https://doi.org/10.1098/rspb.2016.1625
Lundberg SM, Erion G, Chen H et al. (2020) From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2:56–67. https://doi.org/10.1038/s42256-019-0138-9
Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions. In: Advances in neural information processing systems, 30. 4765–4774
Marucco F, Boiani MV, Dupont P et al. (2023) A multidisciplinary approach to estimating wolf population size for long-term conservation. Conservation Biology. https://doi.org/10.1111/cobi.14132
Massolo A, Meriggi A (1998) Factors affecting habitat occupancy by wolves in northern Apennines (northern Italy): a model of habitat suitability. Ecography 21:97–107. https://doi.org/10.1111/j.1600-0587.1998.tb00663.x
Mayer M, Olsen K, Schulz B et al. (2022) Occurrence and livestock depredation patterns by wolves in highly cultivated landscapes. Front Ecol Evol 10: https://doi.org/10.3389/fevo.2022.783027
Mech LD, Boitani L (2003) Wolves: Behavior, Ecology, and Conservation. The University of Chicago Press, Chicago, IL, USA
Mech LD, Dawson DK, Peek JM et al. (1980) Deer Distribution in Relation to Wolf Pack Territory Edges. J Wildl Manag 44:253. https://doi.org/10.2307/3808381
Mech LD, Harper EK (2002) Differential use of a wolf, Canis lupus, pack territory edge and core. Can Field-Naturalist 116:315–316
Mladenoff DJ, Clayton MK, Pratt SD, et al. (2009) Change in Occupied Wolf Habitat in the Northern Great Lakes Region. In: Recovery of Gray Wolves in the Great Lakes Region of the United States. Springer New York, New York, NY, pp 119–138
Mladenoff DJ, Sickley TA, Sickley TA et al. (1995) A regional landscape analysis and prediction of favorable gray wolf habitat in the northern Great Lakes region. Conserv Biol 9:279–294
Modelplotpy (2023) modelplotpy package, 1.0.0. In: https://modelplotpy.readthedocs.io/en/latest/
Nagelkerke J (2022) Visualise the business value of predictive models. In: https://medium.com/p/21c6bc8132c
Naturhistorisk museum Aarhus (2023) Atlas over Danmarks Ulve. In: https://www.ulveatlas.dk/nyheder/ulvehvalpe-i-danmark-for-foerste-gang-i-over-200-aar/
Nowak S, Mysłajek RW, Kłosińska A, Gabryś G (2011) Diet and prey selection of wolves (Canis lupus) recolonising Western and Central Poland. Mamm Biol 76:709–715. https://doi.org/10.1016/j.mambio.2011.06.007
Oakleaf JK, Murray DL, Oakleaf JR et al. (2006) Habitat selection by recolonizing wolves in the northern Rocky Mountains of the United States. J Wildl Manag 70:554–563
Okarma H (1995) The trophic ecology of wolves and their predatory role in ungulate communities of forest ecosystems in Europe. Acta Theriol (Warsz) 40:335–386. https://doi.org/10.4098/AT.arch.95-35
Ordiz A, Uzal A, Milleret C et al. (2020) Wolf habitat selection when sympatric or allopatric with brown bears in Scandinavia. Sci Rep. 10:9941. https://doi.org/10.1038/s41598-020-66626-1
Pasini A (2015) Artificial neural networks for small dataset analysis. J Thorac Dis 7:953–960
Pedregosa F, Varoquaux G, Gramfort A et al. (2011) Scikit-learn: Machine Learning in Python. J Mach Learn Res 12:2825–2830
Reinhardt I, Ansorge H, Collet S et al. (2021) Erkenntnisse zur Wiederausbreitung des Wolfs in Deutschland. 0028-0615 96:19–26. https://doi.org/10.17433/1.2021.50153869.19-26
Reinhardt I, Kluth G (2015) Untersuchungen zum Raum-Zeitverhalten und zur Abwanderung von Wölfen in Sachsen. Endbericht Projekt "Wanderwolf" (2012 - 2014)
Reinhardt I, Kluth G, Jarausch A et al. (2017) Dokumentations- und Beratungsstelle des Bundes zum Thema Wolf. Wölfe in Deutschland - Statusbericht 2015/16.
Reinhardt I, Kluth G, Nowak C et al. (2019) Military training areas facilitate the recolonization of wolves in Germany. Conserv Lett 12:. https://doi.org/10.1111/conl.12635
Ripple WJ, Estes JA, Beschta RL et al. (2014) Status and Ecological Effects of the World's Largest Carnivores. Science (1979) 343:. https://doi.org/10.1126/science.1241484
Roche DG, O'Dea RE, Kerr KA, et al. (2022) Closing the knowledge-action gap in conservation with open science. Conservation Biology 36:. https://doi.org/10.1111/cobi.13835
Roder S, Biollaz F, Mettaz S et al. (2020) Deer density drives habitat use of establishing wolves in the Western European Alps. J Appl Ecol 57:995–1008. https://doi.org/10.1111/1365-2664.13609
Scikit-learn (2023) scikit-learn package 1.2.2. In: https://pypi.org/project/scikit-learn/
SHAP (2023) shap 0.41.0. In: https://pypi.org/project/shap/
Smith JB, Nielsen CK, Hellgren EC (2016) Suitable habitat for recolonizing large carnivores in the midwestern USA. Oryx 50:555–564. https://doi.org/10.1017/S0030605314001227
Stahler D, Heinrich B, Smith D (2002) Common ravens, Corvus corax, preferentially associate with grey wolves, Canis lupus, as a foraging strategy in winter. Anim Behav 64:283–290. https://doi.org/10.1006/anbe.2002.3047
Tuia D, Kellenberger B, Beery S et al. (2022) Perspectives in machine learning for wildlife conservation. Nat Commun 13:792. https://doi.org/10.1038/s41467-022-27980-y
van den Bosch M, Beyer DE, Erb JD et al. (2022) Identifying potential gray wolf habitat and connectivity in the eastern USA. Biol Conserv 273:109708. https://doi.org/10.1016/j.biocon.2022.109708
Van der Veken T, Van den Berge K, Gouwy J et al. (2021) Diet of the first settled wolves (Canis lupus) in Flanders, Belgium. Lutra 64:45–56
van Liere D, Siard N, Martens P, Jordan D (2021) Conflicts with wolves can originate from their parent packs. Animals 11:1801. https://doi.org/10.3390/ani11061801
Wagner C, Holzapfel M, Kluth G et al. (2012) Wolf (Canis lupus) feeding habits during the first eight years of its occurrence in Germany. Mamm Biol 77:196–203. https://doi.org/10.1016/j.mambio.2011.12.004
Witek M, Zwolicki A, Wikar Z et al. (2023) High abundance of an introduced prey species, fallow deer Dama dama, abolishes wolf preference towards red deer. In: Wolves across borders, international conference on wolf ecology and management
XGBoost (2023) xgboost 1.7.5. In: https://pypi.org/project/xgboost/
Zabihi-Seissan S, Prokopenko CM, Vander Wal E (2022) Wolf spatial behavior promotes encounters and kills of abundant prey. Oecologia 200:11–22. https://doi.org/10.1007/s00442-022-05218-4