Received: 17 March 2022 | Revised: 30 November 2022 | Accepted: 23 December 2022

DOI: 10.1111/ddi.13675

RESEARCH ARTICLE

Diversity and Distributions WILEY

Modelling potential biotope composition on a regional scale revealed that climate variables are stronger drivers than soil variables

Sven Rubanschi | **Sebastian T. Meyer** | **Christian Hof** | **Wolfgang W. Weisser**

Terrestrial Ecology Research Group, School of Life Sciences, Technical University Munich, Freising, Germany

Correspondence
Sven Rubanschi, Terrestrial Ecology Research Group, School of Life Sciences, Technical University Munich, D-85354 Freising, Germany.
Email: sven.rubanschi@tum.de

Funding information
Bavarian Ministry of Science and the Arts in the context of the Bavarian Climate Research Network (bayklif, project BLIZ); Deutsche Forschungsgemeinschaft (DFG) through TUM International Graduate School of Science and Engineering (IGSSE), GSC 81

Editor: Martin Jung

Abstract

Aim: Environmental conditions define the suitability of an area for biotopes, and any area can be suitable for several biotopes. However, most previous studies modelled the distribution of single biotopes ignoring the potential co-occurrence of biotopes in one area, which limits the usefulness of such models for conservation and restoration planning. In this study, we described the potential biotope composition of an area in response to environmental conditions.

Location: Bavaria, Federal State of Germany.

Methods: Based on the Bavarian biotope mapping data, we modelled the distribution of 29 terrestrial biotopes based on six climate variables and six chemical and four physical soil properties using the species distribution modelling algorithm Maxent.

Results: For most biotopes, we found that climate variables were more important than soil variables for the biotope distribution and that the area of the predicted biotope distribution was larger than the observed distribution. The potential biotope composition illustrated that while 8% of the area in Bavaria was not sufficiently suitable for any analysed biotope, 92% of the modelled area in Bavaria was suitable for at least one biotope, 84% for two and 77% for at least three biotopes. The difference in suitability between the most suitable biotopes in composition was minor. Further, over one-quarter of the modelled area was suitable for 6–8 different biotopes.

Main Conclusions: Our study showed that considering a composition of potentially suitable biotopes in a raster cell, instead of only the most suitable biotope, provides valuable information to identify conservation priorities and restoration opportunities.

KEYWORDS
Bavaria, biotope distribution, distribution modelling, Maxent, species distribution model

1 | INTRODUCTION

Globally, biodiversity declines, threatening over 82,000 species (Maxwell et al., 2016). Apart from land-use change and the direct exploitation of species, anthropogenic climate change is now among the most important drivers of biodiversity decline (IPBES, 2019; Newbold et al., 2020). It alters the environmental conditions to such an extent that many ecoregions will be put under substantial survival stress (Beaumont et al., 2011), forcing them to adapt their environmental demands (Sillero et al., 2022) or to move to more suitable areas, which poses multiple challenges to them. It is unclear whether they can arrive at a new area with suitable conditions due to landscape destruction and fragmentation (Hof, 2021), and even if a species can reach a new area with suitable environmental conditions, it is questionable if local human activities allow the occurrence of the species (cf. effects of land use; Franklin, 1995).

Measures to reduce biodiversity decline need to be accompanied by biodiversity monitoring, but monitoring data are available for only a few well-studied species groups in a few well-sampled regions (IUCN, 2021). Comprehensive monitoring of all species will take too long, but conservation planning needs information now. One way to overcome this challenge is to monitor and preserve whole biotopes instead of single species (Chytrý et al., 2020; European Community, 1992; Watson et al., 2014). Biotopes are landscape elements characterised by specific species communities (primarily plant species composition) that developed due to environmental conditions (Colwell & Rangel, 2009). Thus, biotopes represent specific land-cover types that can be displaced by other land-cover types like agricultural fields. Monitoring and conserving biotopes is assumed to monitor and conserve the associated species and biodiversity (Watson et al., 2014), thereby mitigating the challenge of lacking biodiversity data.

For some biotopes (and their associated species), the currently remaining areas might not be large enough to ensure long-term survival due to insufficient population sizes (Matthies et al., 2004). Future climate and land-use change will further exacerbate this problem (Hof et al., 2011). Therefore, besides conservation, active biotope restoration is needed. To select areas suitable for restoration, information about the potential distribution of biotopes is crucial (Loidi & Fernández-González, 2012). Rather than depending on qualitative expert opinions, which was classically done in potential natural vegetation approaches (Tüxen, 1956), information on potential biotope distribution can be provided quantitatively by distribution modelling (i.e. the predicted distribution of biotopes, Guisan et al., 2017; Jiménez-Alfaro et al., 2018). These models evaluate the abiotic environmental conditions of an area and estimate the potential suitability for a biotope based on the concept that abiotic environmental variables describe the potential distribution of a biotope (Franklin, 1995). Ideally, this potential distribution would be available for each biotope so that conservation and restoration could consider multiple biotopes. However, the focus of previous studies was only on a few specific biotopes with high conservation interest (e.g. Keith et al., 2014; Marage & Gégout, 2009; Vogiatzakis & Griffiths, 2006), or the studies were on coarse spatial scales (Jiménez-Alfaro et al., 2018), which limits the use for regional management and planning (Rubanschi et al., 2022).

In this paper, we model the potential biotope composition by considering multiple environmentally suitable biotopes in one area. This approach will provide crucial information for conservation and restoration planning. We use the unique dataset of the Bavarian biotope mapping (Lang & Zintl, 2018) that describes the occurrence of an extensive range of different biotopes in the German Federal State of Bavaria. For a selection of 29 terrestrial biotopes, we project the environmental suitability in Bavaria based on 16 abiotic variables covering climate and soil properties using the Maxent species distribution modelling algorithm. Based on these projections, we identified the potential biotope composition for the whole of Bavaria. Specifically, our study aimed at (a) identifying the environmental variables that drive the distribution of single biotopes, (b) illustrating which biotopes can potentially co-occur in an area and (c) showing how many biotopes can potentially co-occur.

2 | METHODS

2.1 | Study region

The study was conducted in the Federal State of Bavaria (southeast Germany), covering an area of around 70,550 km^2 located between 47°16′-50°34′N and 8°58′-13°50′E. The Bavarian elevation profile includes the Calcareous Alps in the south (2962 m a.s.l. on Mt. Zugspitze), the Bavarian Forest in the north-east (1455 m a.s.l. on Mt. Arber), the hook-shaped Franconian Jura in the centre (600–700 m a.s.l.) and the lowlands between 100 and 500 m a.s.l. The Bavarian climate ranges from sub-oceanic in the northwest to subcontinental in the plains and basins, to a montane climate in the Alps. Bedrock also varies from granite and gneiss in the Bavarian Forest to limestone in the Alps and the Franconian Jura. Land use in Bavaria is dominated by agriculture (46.3%) and forestry (35.3%) (Bayrisches Landesamt für Statistik, 2020).

2.2 | Environmental conditions in Bavaria and their selection

To describe the current environmental conditions across Bavaria, we collected 19 climate variables from the WorldClim dataset (Version 2.1, spatial resolution of 30 arc seconds, temporal aggregation of the years 1970–2000; Fick & Hijmans, 2017), nine soil chemical properties (spatial resolution of 250 m; Ballabio et al., 2019), six soil physical properties (spatial resolution of 250 m; Ballabio et al., 2016) and elevation as a topographical variable (spatial resolution of 30 arc seconds; European Environment Agency, 2016), resulting in a total of 35 variables that could be used for predicting the biotope distribution (Table 1).

The environmental variables were provided in different spatial resolutions (see Table 1), which we resolved by using the coarsest grid to avoid the need for downscaling variables. We thus created a raster based on the resolution of the climatic variables (30 arc seconds, 56.6 ha $\pm$ 0.9 ha, i.e. a square of c. 930 m $\times$ 610 m in the study region) and rescaled the other environmental variables by calculating for the new raster cells the mean value of the initial raster cells

TABLE 1 Summary of environmental variables and their spatial resolution

Abiotic variables (spatial resolution)	Short name	Unit
Climate (30 arc seconds)		
Annual Mean Temperature		°C
Mean Diurnal Range		°C
Isothermality	Isothermality	%
Temperature Seasonality	Temp_Seasonality	°C × 100
Max Temperature of Warmest Month		°C
Min Temperature of Coldest Month		°C
Temperature Annual Range		°C
Mean Temperature of Wettest Quarter	Mean_Temp_Wettest_Quarter	°C
Mean Temperature of Driest Quarter	Mean_Temp_Driest_Quarter	°C
Mean Temperature of Warmest Quarter		°C
Mean Temperature of Coldest Quarter		°C
Annual Precipitation		mm
Precipitation of Wettest Month		mm
Precipitation of Driest Month		mm
Precipitation Seasonality	Prec_Seasonality	%
Precipitation of Wettest Quarter		mm
Precipitation of Driest Quarter		mm
Precipitation of Warmest Quarter		mm
Precipitation of Coldest Quarter	Prec_Coldest_Quarter	mm
Soil chemical properties (250 m)		
Cation Exchange Capacity	Cation_Exchange_Capacity	cmol × kg^{-1}
C/N ratio		ratio
Calcium carbonates	$CaCO_3$	g kg^{-1}
Nitrogen	Nitrogen	g × kg^{-1}
pH in $CaCl_2$		pH
pH in H_2O		pH
pH in H_2O minus pH in $CaCl_2$	pH_H2O_CaCl_ratio	pH
Phosphorus	Phosphorus	mg × kg^{-1}
Potassium	Potassium	mg × kg^{-1}
Soil physical properties (250 m)		
Available Water Capacity	Available_Water_Capacity	mm/m
Bulk density	Bulk_density	T × m^{-3}
Clay content		%
Coarse fragments	Coarse_fragments	%
Sand content		%
Silt content	Silt%	%
Topology (30 arc seconds)		
Elevation		m

Note: Variables used in the Maxent models are italic and have a short name, which is used in Figure 1.

overlapping with the new raster cells. This rescaling was performed with the geographic information system QGIS (QGIS Development Team, 2020). 385 out of a total of 126,697 raster cells were excluded as they were entirely covered by water bodies or had a high proportion of sealed surfaces such as airports or cities.

To correctly assess the association of biotope distribution with environmental variables, the variables require low collinearity and low variance inflation factors (VIFs; Dormann et al., 2013; Zuur et al., 2010). Therefore, we excluded variables that correlated with $|r| > .7$ (Figure S1 in Appendix S3; Dormann et al., 2013) or had a VIF > 3 (Table S6 in Appendix S2; Zuur et al., 2010). After the stepwise elimination of variables using the “vifcor” and “vifstep” functions of the “usdm” packaged 1.1-18 (Naimi et al., 2014) in R 3.6.1 (R Core Team, 2020), 16 variables remained, including six climate variables, six chemical and four physical soil properties (Table 1). The variable selection eliminated the topographical variable elevation because of its collinearity with climate variables (Figure S1 in Appendix S3).

2.3 | The Bavarian biotope mapping

Since 1985, the Bavarian Environment Agency (Bayerisches Landesamt für Umwelt) maps biotopes by monitoring the entire area of Bavaria. They aim to map all (semi-)natural areas housing specific biological communities containing often protected or threatened species (Rubanschi et al., 2022). Consequently, most mapped biotopes are protected under federal and state nature conservation acts (§ 30 and 39 of the BNatSchG/Federal Nature Conservation Act, articles 16 and 23 of the BayNatSchG/Bavarian Nature Conservation Act). For each mapped biotope, the spatial shape of the area (polygon) and vegetation composition were recorded. This vegetation composition was then compared with a classification key (Lang & Zintl, 2018) to assign the observation to one of 108 biotopes (Table S1 in Appendix S2). Since 2006, some of the 108 biotopes were even further differentiated into subtypes according to the Fauna-Flora-Habitat (FFH) guidelines, which aim to protect wild species and their habitats and to provide a European-wide network of habitats (European Community, 1992; Lang & Zintl, 2018). Due to small-scale fluctuation in species composition, many recorded polygons consisted of mixtures of different biotopes (e.g. 10% "Nutrient-poor old grasslands and fallow grasslands", 20% "Nutrient-poor grasslands, base-rich" and 70% "Hedges, near natural"). In total, 1.7 million biotopes were observed, covering about 5% (3723 km^2) of the area of Bavaria (Rubanschi et al., 2022).

2.4 | Preparation of biotope data

Using the Bavarian biotope mapping dataset, we treated each biotope per polygon as an independent observation of this biotope. Further, we focused on the 108 main biotopes and ignored the additional differentiation into subtypes. Instead, we grouped the biotopes into six biotope types based on the dominating vegetation described in the biotope mapping manual (Lang & Zintl, 2018), covering the following proportions of the total mapped biotope area: bush (11.4%), forest (23.5%), grassland (32.6%), peatland (13.9%), water-associated (9.5%) and biotopes, which were human-dominated or had no clear definition (9.1%) (see Appendix 1 and Rubanschi et al., 2022).

For the biotope distribution modelling, we focused on terrestrial biotopes (bushes, forests and grasslands) that provided a sufficient number of observations. Thus, we excluded all peatland, water-associated and human-dominated biotopes and those biotopes with <500 observations, resulting in 29 biotopes with a total number of 685,647 observations (Table 2; 39.5% of all biotope observations) covering 2028 km^2 (54.5% of the mapped biotope area) with an average polygon size of 0.51 ha.

Abiotic variables were arranged as raster cells and biotopes were recorded as polygons. Most of the biotope polygons (71.05%) were located in single raster cells that represent the environmental conditions of that polygon. The remaining 28.95% of biotope polygons extended over multiple raster cells. To describe the environmental conditions of these polygons, we calculated for each observed polygon the weighted mean of environmental condition (WM, see Equation 1) by multiplying the environmental value (EV, see Equation 1) of a raster cell overlapped by the polygon with the proportion of the whole polygon in the respective raster cell ($POLY_{proportion}/POLY_{whole}$, see Equation 1). Then, we sum the weighted mean values over all raster cells overlapped by one polygon (Equation 1, n equals the number of raster cells a polygon is overlapping with, i equals the raster cell in which the respective proportion of polygon overlapped).

$$WM = \sum_{i=1}^{n} EV_i * \frac{POLY_{proportion_i}}{POLY_{whole}} \quad (1)$$

2.5 | Modelling and evaluating the potential distribution of biotopes

All biotopes occurring in Bavaria are recorded in the Bavarian biotope mapping as spatially explicit polygons. However, biotopes are land-cover types that can be displaced by anthropogenic land use. Thus, the absence of biotopes in an area cannot be interpreted as evidence for unsuitable abiotic conditions. Therefore, we treat the presence/absence dataset of the biotope mapping as a presence-only dataset (cf., Elith et al., 2020; Lobo et al., 2010). We used the observations of a biotope in a polygon and the weighted mean environmental conditions of the respective polygon as presence data. Raster cells that do not overlap with the respective polygons were considered as background data in the model. The majority of

TABLE 2 Summary of selected biotopes showing their biotope type and number of observations

Biotope	Class	Observations
Alluvial forest	Forest	21,950
Alpine lawn	Grassland	5319
Alpine nutrient-poor grassland	Grassland	1799
Alpine yellow oat grasslands	Grassland	621
Beech forest, heat-loving	Forest	946
Black alder forest	Forest	1015
Block and rubble forests	Forest	583
Broadleaf forest, mesophilic	Forest	11,522
Broadleaf forest, soil acidic	Forest	2241
Dwarf shrubs and gorse heath	Bushes	3784
Heat-loving edges	Grassland	7873
Heat-loving shrubs	Bushes	2651
Hedges, near natural	Bushes	197,402
Large sedge meadows outside the siltation zone	Grassland	16,161
Mesophilic shrubs, near natural	Bushes	56,741
Moist and wet tall herbaceous vegetation, planar to montane	Grassland	56,664
Mountain pine scrubs	Bushes	2237
Nard grass lawn	Grassland	6617
Nutrient-poor grasslands, base-rich	Grassland	44,552
Nutrient-poor old grasslands and fallow grasslands	Grassland	89,116
Nutrient-poor yellow oat grasslands	Grassland	645
Pine forests, base-rich	Forest	1536
Pine forests, soil acidic	Forest	709
Pipegrass meadows	Grassland	7133
Sandy nutrient-poor grasslands	Grassland	3677
Species-rich extensive grassland	Grassland	66,442
Species-rich lowland meadows of medium sites	Grassland	6085
Watercourse companion trees, linear	Forest	54,812
Wetland shrubs	Bushes	14,814

biotope polygons were observed in different raster cells (<3% of biotope polygons occurred in raster cells where another polygon of the same biotope was observed), and on average, the distance between the observed biotope polygons was high (mean distance 62.35–209.98 km, median distance 24.45–234.86 km for individual biotopes). Consequently, we think spatial autocorrelation to be a minor issue with this data and saw no requirement for specific algorithms, which correct for spatial autocorrelation. We modelled the distribution for each of the 29 biotopes in response to environmental conditions in a separate model (see ODMAP protocol in Appendix S1; Zurell et al., 2020).

We used the algorithm Maximum Entropy (Maxent) to model the distribution of the biotopes (Phillips et al., 2006). Maxent has already been used to predict the distribution of vegetation communities (Fischer et al., 2019; Hemsing & Bryn, 2012; Jiménez-Alfaro et al., 2018; Tarkesh & Jetschke, 2012) and typically showed higher predictive accuracy than other distribution modelling approaches (Merow et al., 2013; Warren et al., 2020). Here, Maxent predicts a biotope's distribution by finding its distribution with maximum entropy (i.e. that is closest to uniform), where the average value of each environmental variable for the predicted distribution equals the average of the variable in the observed distribution of the biotope (Phillips et al., 2006). Thus, we infer the predicted distribution from the environmental conditions at the observations without placing any unjustified constraints (Tarkesh & Jetschke, 2012). This also means that Maxent does not necessarily consider the absence of a biotope as evidence for unsuitable environmental conditions (see Phillips et al., 2006), which encourages our interpretation of the dataset. Since we used Maxent with a present-only dataset, the model output represents the relative probability of a biotope being present in a raster cell (Halvorsen, 2012). Often this probability is compared with the suitability of a raster cell, which is calculated with a present/absent dataset. To simplify the interpretation of the model predictions, we considered the relative probability of presence as suitability since the predictions of present-only and present/absent datasets are comparable (Hirzel et al., 2006).

The biotope distribution models were calculated with Maxent version 3.4.1 (Phillips et al., 2006) using the package "dismo" 1.1-4 (Hijmans et al., 2017) in R 3.6.1 (R Core Team, 2020). Maxent settings were: maximum number of iterations = 10,000, convergence threshold = 0.00001 and model output = logistic. The rest of the model settings were set to default (allowing all feature transformation; see Table S5 in Appendix S2). To assess the variability in model accuracy, we used a bootstrap approach with 100 repetitions per model (Efron, 1992). For each repetition, Maxent used 90% of the data to fit the model (training data) and 10% for later evaluation (test data). After fitting the distribution model based on the training data, we predicted the suitability for the reminding 10% of test data, which the model had not seen. Then, we calculated the area under the receiver-operator curve (AUC; Mason & Graham, 2002) for the test dataset. The AUC measures the probability that the model assigns higher suitability to a random presence observation than a random background point (Liu et al., 2011; Phillips et al., 2006). In other words, AUC is high when differences in predicted suitability between present observations and the background points of a model are large. Unlike other performance metrics, it is independent of a preselected suitability threshold and can be applied directly to the row model output (Liu et al., 2011). AUC values can range from 0 to 1, where 1 indicates a model that perfectly predicts the distribution of the biotope according to its observation, while 0.5 is a model that is not better than random guesses. Generally, the predictability of the models can be considered reasonable when the AUC values are over 0.7 (Pearce & Ferrier, 2000; Swets, 1988). Because of the presence-only approach of the Maxent algorithm, the maximum achievable AUC is slightly lowered to <1 (Phillips et al., 2006). Although our main focus was on the threshold-independent performance metric AUC we also calculated the true skill statistic (TSS), which is a threshold-dependent performance metric that ranges between −1 and +1, where +1 indicates that the model perfectly predicts the biotope's distribution according to its observation, while 0 indicates a model that is not better performing than random guesses (Allouche et al., 2006).

To summarise the results of the separate bootstrap runs, we calculated the mean and standard deviation of AUC for each biotope distribution model (Table 3). Then, we predicted with each of the 100 models per biotope (bootstrap repletion) the suitability of each Bavarian raster cell for this biotope by calculating the mean suitability across all models resulting in a suitability map for each biotope in Bavaria (Figure S1 in Appendix S2). To compare the predictions of the different models, we selected the logistic format, which is monotonically related to the raw Maxent output (Elith et al., 2011). With that, the suitability values ranged between 0 and 1, where 0 indicates that the environmental conditions were unsuitable and 1 perfectly suitable for the biotope.

To test the robustness of our results, we compared the results of the Maxent models to results obtained with generalised adaptive models (GAM; Guisan et al., 2017) using the same presence and background data as for the Maxent models (just the number of iterations was reduced to 5). Results of the GAMs are presented as sensitivity analyses (Figure S4 and Table S3 in Appendix S2).

2.6 | Importance of environmental variables for the biotope distribution

To evaluate which environmental conditions affected the distribution of individual biotopes, we calculated the permutated variable importance (Phillips, 2017). There, one variable after the other was randomly permuted over all raster cells. As a result, there was a loss in accuracy, which was recorded and transformed into a percentage value for each variable. This value represents the importance of the variable for the predictive accuracy of the model. In other words, a large accuracy loss due to the permutation of the variable indicates that the distribution of the biotope depends strongly on that variable, which is characterised by a high percentage (Phillips, 2017). Per biotope and each bootstrap repetition, the model yielded a permutated importance for each environmental variable. To summarise the permutated importance over all bootstrap repetitions, we calculated

TABLE 3 Summary of the biotope model mean AUCs and their standard deviation (SD)

Biotope	AUC	SD
Alluvial forest	0.73	0.00
Alpine lawn	0.98	0.00
Alpine nutrient-poor grassland	0.99	0.00
Alpine yellow oat grasslands	0.99	0.00
Beech forest, heat-loving	0.95	0.01
Black alder forest	0.83	0.02
Block and rubble forests	0.99	0.00
Broadleaf forest, mesophilic	0.78	0.01
Broadleaf forest, soil acidic	0.85	0.01
Dwarf shrubs and gorse heath	0.89	0.01
Heat-loving edges	0.88	0.01
Heat-loving shrubs	0.90	0.01
Hedges, near natural	0.59	0.00
Large sedge meadows outside the siltation zone	0.70	0.01
Mesophilic shrubs, near natural	0.75	0.00
Moist and wet tall herbaceous vegetation, planar to montane	0.67	0.00
Mountain pine scrubs	0.99	0.00
Nard grass lawn	0.91	0.00
Nutrient-poor grasslands, base-rich	0.78	0.00
Nutrient-poor old grasslands and fallow grasslands	0.67	0.00
Nutrient-poor yellow oat grasslands	0.99	0.00
Pine forests, base-rich	0.94	0.01
Pine forests, soil acidic	0.96	0.01
Pipegrass meadows	0.89	0.01
Sandy nutrient-poor grasslands	0.86	0.01
Species-rich extensive grassland	0.68	0.00
Species-rich lowland meadows of medium sites	0.86	0.01
Watercourse companion trees, linear	0.66	0.00
Wetland shrubs	0.74	0.01

its mean and standard deviation per environmental variable and biotope (Table S2 in Appendix S2).

2.7 | Potential biotope composition in Bavaria

For the 29 biotopes, the models provided a suitability value for each raster cell. Converting a continuous suitability value to a presence-absence map requires applying a suitability threshold. Commonly, a biotope-specific threshold is selected that minimises the discrepancy between predicted and observed distribution (Liu et al., 2005). To avoid using a threshold that depends on the observed biotope distribution (as this observed distribution is known to be strongly affected by anthropogenic land use), we interpreted the suitability value in its original sense as a value that evaluates how suitable the abiotic environmental conditions in a raster cell are for a particular biotope. This way, we selected a suitability threshold of 0.5 to differentiate between suitable or unsuitable abiotic environmental conditions for the biotope in a raster cell (cf. Bailey et al., 2002; Stockwell & Peterson, 2002; Woolf et al., 2002). We are aware that suitability values are affected by prevalence, which is the proportion between biotope observation and background points (Santika, 2011). Biotopes with high prevalence will tend to have generally higher suitability values, while biotopes with low prevalence will tend to have generally low suitability values (Jiménez-Valverde & Lobo, 2007; Santika, 2011). However, since the prevalence in the modelling process (training samples/background points), was lower than 0.5 (Table S1 in Appendix S3) for all biotopes except for "Hedges, near natural", we expected that the selected suitability threshold of 0.5 will not inflate the potential occurrence of biotopes but rather evaluate the potential occurrence of biotopes with low prevalence more conservatively. To assess the effects of prevalence and the sensitivity of the results to the chosen threshold value, we repeated the analyses with thresholds of 0, 0.25 and 0.75 (Figures S2 and S3 in Appendix S2).

To characterise the potential biotope composition, we counted the number of biotopes that can occur in a raster cell ("potential biotope richness" hereafter) and visualised this number in a map (Figure 2). If all biotopes were excluded from a raster cell due to suitability values below the threshold, these raster cells would be considered unsuitable for any of the biotopes and the potential biotope richness would be zero. For each biotope, we calculated the prevalence of the biotope depending on the potential biotope richness in the raster cells (Figure 3).

For visualising the three most suitable biotopes in a raster cell, we produced maps that show either the biotope with the highest, 2nd highest or 3rd highest suitability (Figure 4). If a raster cell contained fewer than three biotopes, due to the applied threshold, we considered these raster cells in the respective suitability order as unsuitable.

3 | RESULTS

3.1 | Accuracy of the biotope distribution models

The AUC values for 24 of the 29 modelled biotopes were over 0.7 (Table 3) and are conceded as reasonable (Pearce & Ferrier, 2000; Swets, 1988). Alpine biotopes like "Alpine lawn" showed the highest AUC values with 0.99 while "Hedges, near natural" showed the lowest AUC value with 0.59.

3.2 | Importance of environmental variables for the biotope distribution

Generally, climatic variables were more important for the biotope distribution than soil chemical or soil physical properties (Figure 1).

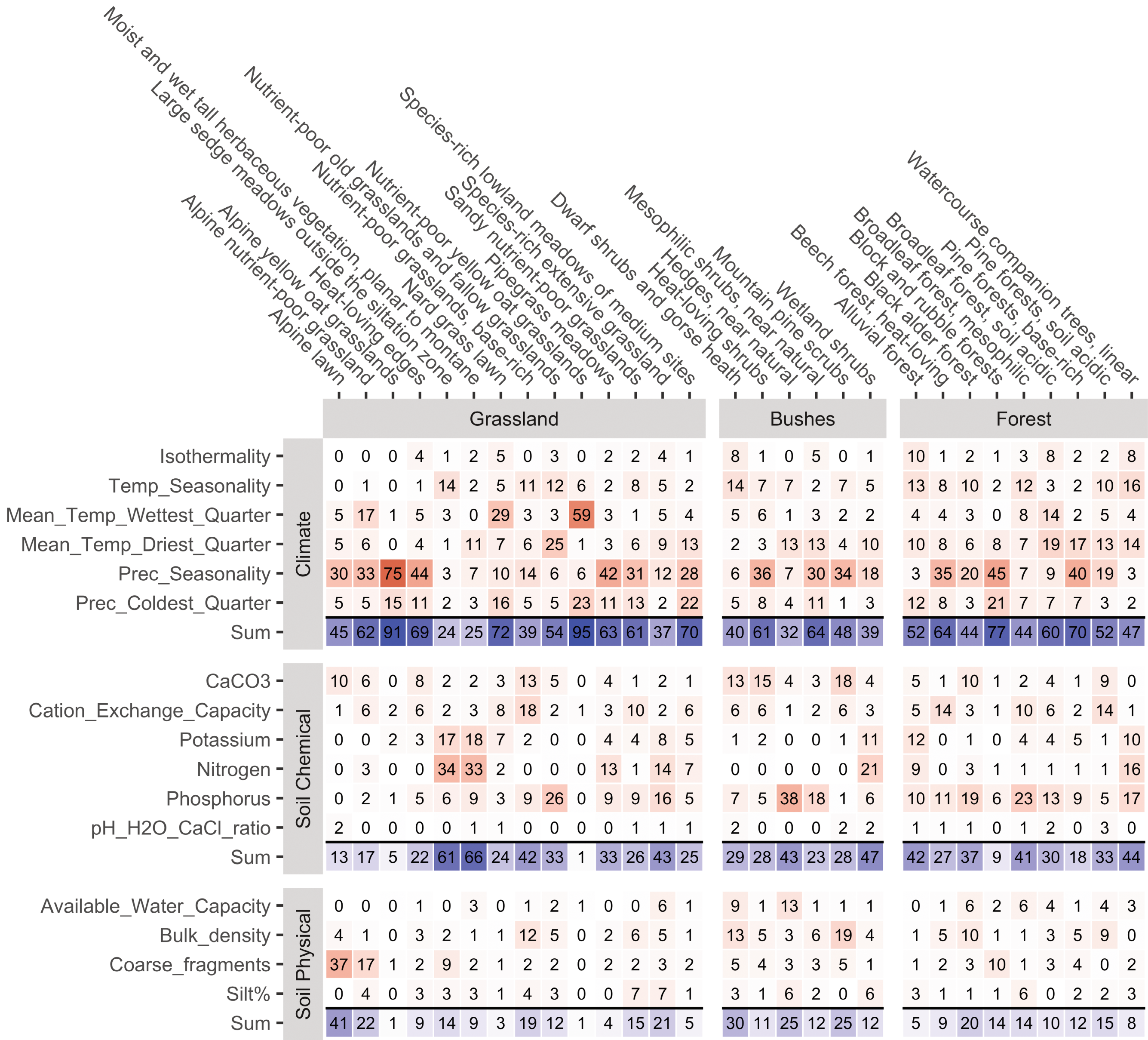

FIGURE 1 Results of Maxent model showing the average importance scores in percentage of the 16 climate and soil variables for the 29 biotopes. The full names of environmental variables can be found in Table 1. Biotopes and environmental variables are grouped by their types. The sum of the importance of each environmental variable group per biotope is provided under the black line. All values are rounded and are reflected by the red colour's intensity for the individual values and the blue colour's intensity for the summed values.

For most of the biotopes, *precipitation seasonality* was the most important variable (Figure 1). This is especially true for "Alpine yellow oat grasslands", for which *precipitation seasonality* had a variable importance of 75%. Other variables that strongly influenced the occurrence of biotopes (importance score of up to 25%) were *mean temperature of wettest quarter, mean temperature of driest quarter* and *precipitation of coldest quarter.*

While soil variables were generally seldom selected as important variables, there were some exceptions. For example, for "Moist and wet tall herbaceous vegetation, planar to montane" and "Large sedge meadows outside the siltation zone", *nitrogen* and *potassium* played an important role. *Phosphorus* was important for biotopes like "Hedges, near natural" and "Nutrient-poor old grasslands and fallow grasslands"; however, the AUC value for these biotopes was low. Additionally, the models of "Alpine lawn" and "Alpine nutrient-poor grassland" rated the soil physical property *coarse fragments* as important.

3.3 | Assessing the potential biotope composition in Bavaria

For most biotopes, the potential distribution was larger than the observed distribution (Figure S1 in Appendix S2). Further, the average

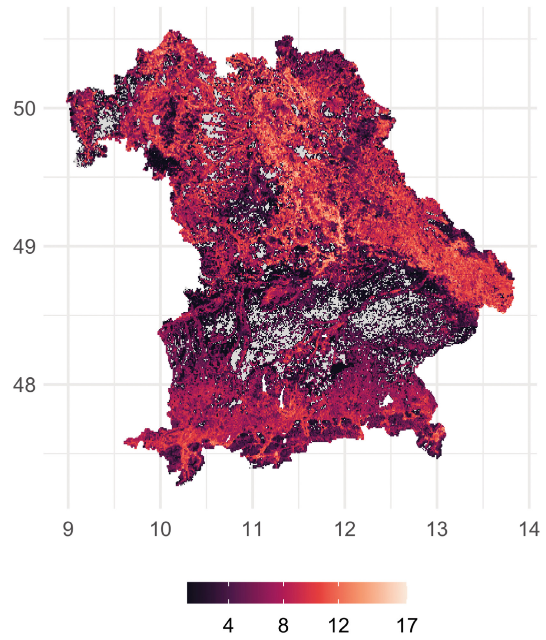

FIGURE 2 Number of biotopes that were considered suitable in a raster cell. Raster cells shown in grey do not feature a biotope with a suitability value >0.5. See text for explanations.

suitability value of the most suitable biotope in a raster cell was 0.813. Comparing the suitability values in raster cells with biotope compositions of three biotopes showed a difference between the first and second most suitable biotope of 0.073, on average, and 0.122 between the first and third, on average.

The highest predicted potential biotope richness was 17, found in 10 raster cells (0.01% of all raster cells), which was slightly higher than the highest observed biotope richness of 14. Most raster cells showed a potential biotope richness of 6–8 different biotopes (Figure 3), totalling one-quarter of all raster cells. Raster cells with high potential biotope richness were found in the Franconian Jura, Bavarian Forest and the pre-alpine lands in the south (Figure 2). Further, a high number of raster cells provided suitable conditions for only one biotope (Figure 3, 10,653 raster cells, 7.78% of the raster cells).

Some biotopes like “Hedges, near natural” or “Large sedge meadows outside the siltation zone” showed generally high prevalence in all potential biotope compositions (Figure 3). With an increasing number of suitable biotopes for a raster cell, we saw a homogenisation in biotope composition (Figure 3). Only a few other biotopes, with generally rare prevalence, occurred in those biotope compositions.

With the application of the 0.5 threshold, 92% of Bavaria was considered suitable for at least one biotope, 84% for two and 77% for up to three biotopes, which usually belonged to different biotope types. The remaining 8% of Bavaria showed unsuitable conditions for any of the analysed biotopes. The map illustrating the biotopes with the highest suitability in a raster cell (Figure 4a) showed that one biotope often dominated large regions of Bavaria. In other words, many neighbouring raster cells provided the most suitable conditions for the same biotope. Some of these regions were also dominated by one biotope in the map illustrating the biotopes with the second-highest suitability in a raster cell but often from a different biotope type (Figure 4b). This pattern ended in a rather noisy map of different neighbouring biotopes in the map illustrating the biotopes with the third-highest suitability in a raster cell (Figure 4c).

The sensitivity analyses (Figures S2 and S3 in Appendix S2) showed that the number of suitable biotopes in a raster cell decreased with a higher suitability threshold. Further, the difference in suitability values of the most suitable biotopes decreased by applying higher thresholds (Table S2 in Appendix S3). Thus, the results became more conservative with higher thresholds since the raster cells were only considered suitable for particular biotopes with very high suitability values. However, the main results that most raster cells are suitable for several biotopes and the regions that provide suitable conditions for a high number of biotopes remained the same (Figures S2 and S3 in Appendix S2). The comparison between Maxent models and the GAMs showed a strong correlation between the model predictions and similarities in model accuracy (Figure S4 and Table S3 in Appendix S2).

4 | DISCUSSION

We found that climate variables were more important than soil variables in driving the biotope distribution for most biotopes. Based on the knowledge of which variables were driving the biotope distribution, we projected the potential distribution of biotopes. The suitable area (potential distribution) for most biotopes was larger than the observed distribution. Merging the predictions for individual biotopes into a potential biotope composition showed that the environmental conditions in more than three-quarters of Bavaria were considered suitable for at least three biotopes. Further, over one-quarter of Bavaria was considered suitable for 6–8 different biotopes. However, our results also indicate that environmental conditions in 8% of Bavaria may be unsuitable for any biotope (suitability values below 0.5).

4.1 | Caveats of modelling the potential distribution of biotopes

Before covering the effects of environmental variables on the potential biotope distribution and composition, questions about model input, model selection and the usage of a threshold need to be discussed (Guisan et al., 2017). In this study, we analysed data from a monitoring whose goal was to describe the presence and absence of all biotopes in Bavaria (Lang & Zintl, 2018). Since, previous studies (IPBES, 2019; Newbold et al., 2020) showed that the occurrence of biotopes is strongly influenced by anthropogenic land use the absence of a biotope is not necessarily the evidence for unsuitable environmental conditions. Therefore, we treated the dataset as a presence-only instead of a presence/absence dataset.

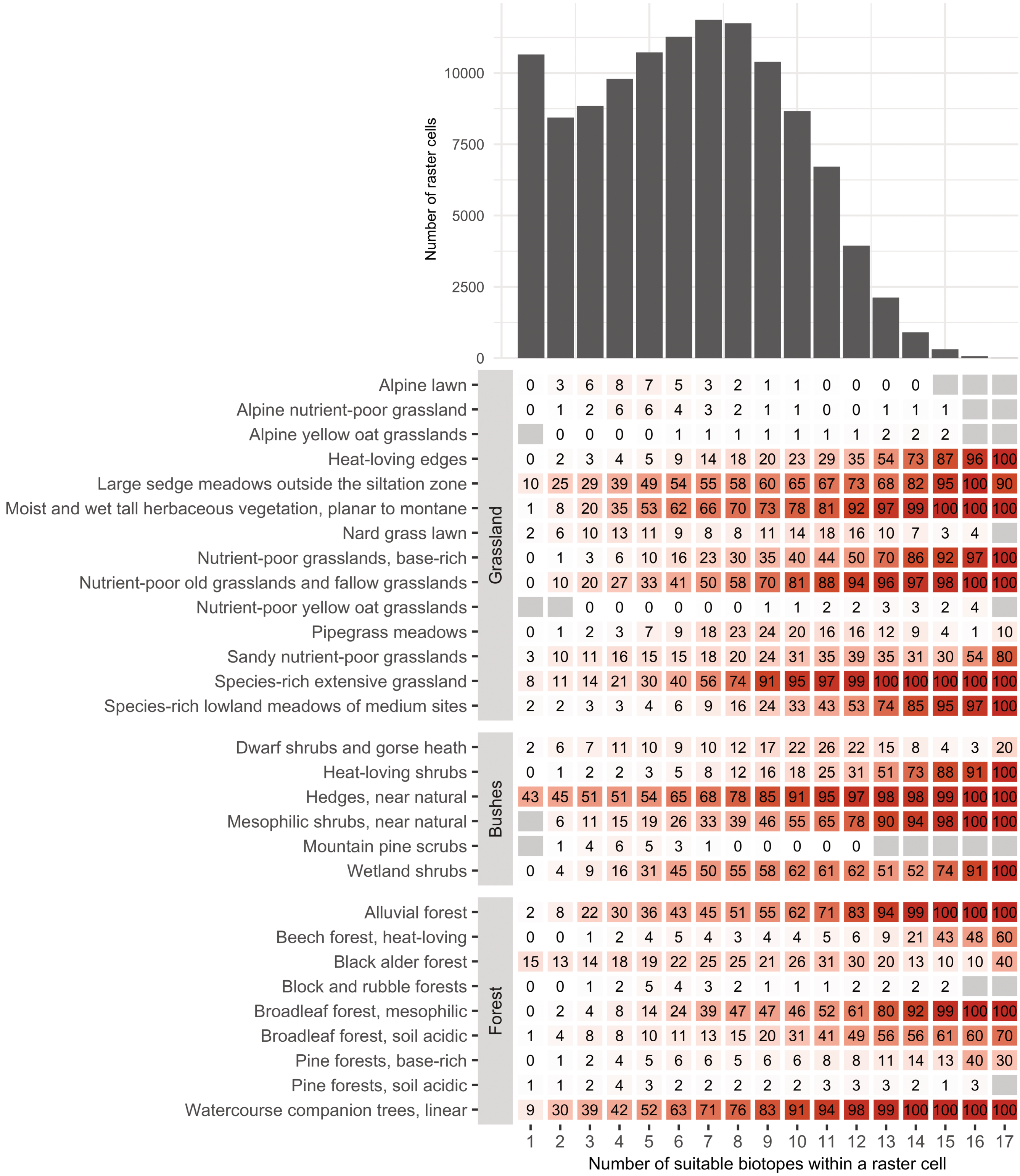

FIGURE 3 Number of raster cells providing suitable conditions for a certain number of biotopes (upper histogram). Values in the lower heatmap explain the prevalence of a biotope in raster cells that provide suitable conditions for a certain number of biotopes. No value and a grey panel in the heatmap indicate that the biotope was considered unsuitable in any raster cell of the raster cells providing suitable conditions for the respective number of biotopes. All values are rounded and are reflected by the red colour's intensity.

Maxent is an established, powerful algorithm to analyse distributions based on presence-only datasets (Merow et al., 2013; Warren et al., 2020). Nevertheless, to show the consistency of the Maxent predictions, they need to be compared with predictions of other distribution models. Here, we compared them with the predictions of GAMs and saw a strong correlation between them and similarity in

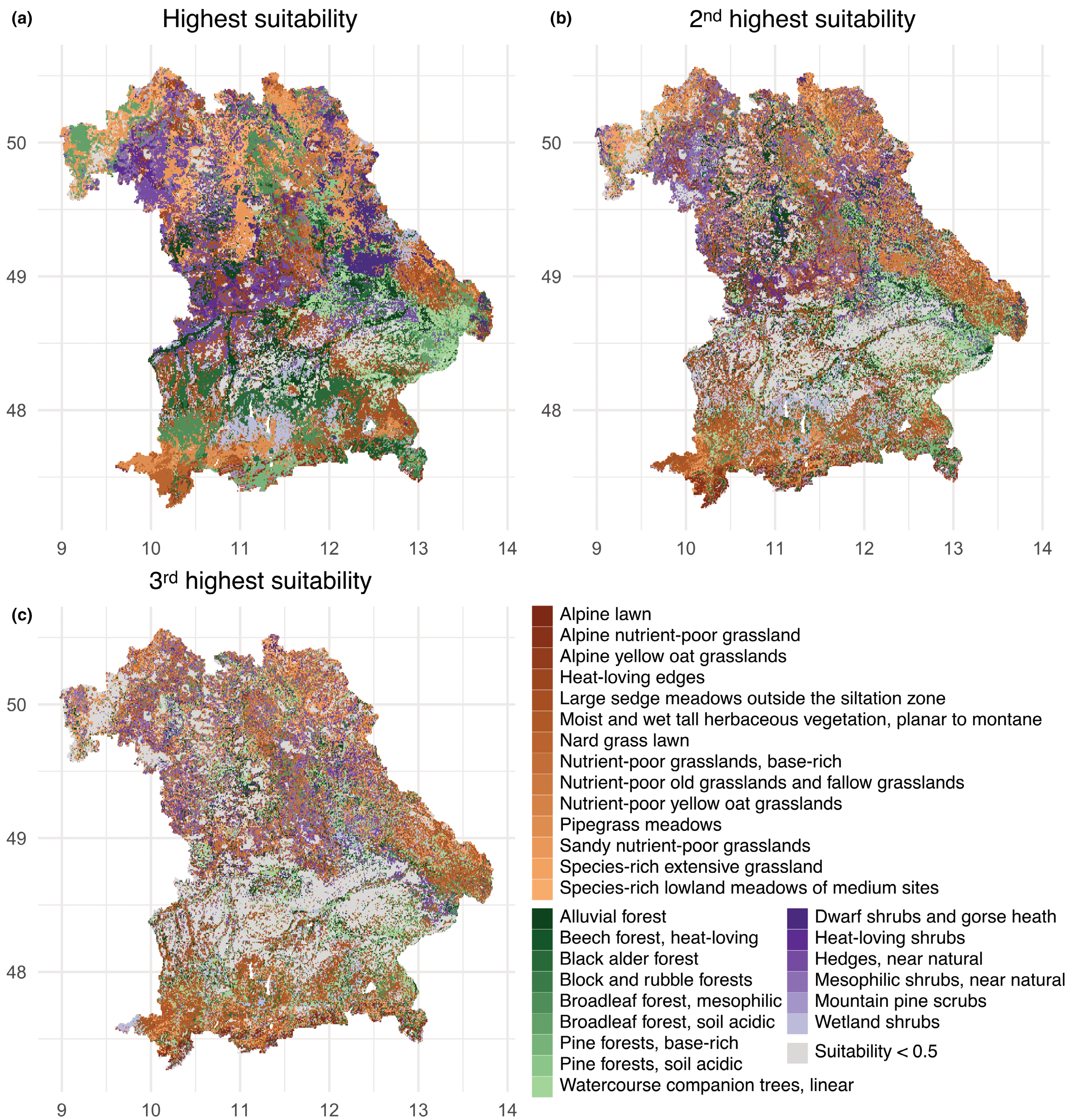

FIGURE 4 Predicted biotope distributions in Bavaria. The three maps show for each raster cell the biotope with the (a) highest, (b) 2nd highest and (c) 3rd highest suitability. Raster cells shown in grey do not feature a biotope with a suitability value >0.5. See text for explanations.

model accuracy (Figure S4 and Table S3 in Appendix S2), confirming the validity of the Maxent predictions. However, the Maxent models provided better interpretable predictions than the GAMs. Maxent predicted values were evenly distributed between 0 and 1, while some GAMs predicted only values between 0 and 0.5 or less (e.g. see "Black alder forest", Figure S4 in Appendix S2). Consequently, GAMs depended on the selection of very low thresholds to provide reasonably predicted distributions. Because these GAM predictions are based on the observation data, the models minimise the divergence between predicted and observed distribution (thus, including the effects of land-use and environmental conditions). Consequently, predicted areas of suitable environmental conditions were underestimated. By contrast, Maxent models considered raster cells without a biotope observation not necessarily unsuitable for the biotope concerning the environmental conditions (see Phillips et al., 2006). Therefore, we consider Maxent predictions more useful because they provide an intuitive interpretation and are not as sensitive to land use affecting the model's predictions. Other approaches like stacked or joint distribution models (Zurell et al., 2020) that also consider biotic interactions could be promising future research avenues. Typically, these interactions are modelled between single species, which might be more challenging when considering whole biotopes with their associated species compositions.

Regarding model performance indicators, we compared the threshold-independent AUC to the threshold-dependent TSS, which indicated similar model accuracy (Table S3 in Appendix S2). In contrast to AUC, TSS depends on the present biotope distribution, which may be strongly affected by anthropogenic land use. The thresholds selected to calculate the TSS were generally lower than the 0.5 threshold that we used to translate Maxent suitability to potential biotope occurrences (Table S4 in Appendix S2). Due to the lower threshold selected to calculate TSS, the number of true positive predictions increased, however, introducing at the same time a high number of false positive predictions (commission error) due to many biotopes that had few observations compared with many background points (Table S1 in Appendix S3; see Allouche et al., 2006). However, all additional (true and false) positive predictions due to thresholds lower than 0.5 were in raster cells with low suitability for the biotope. Thus, models would recommend less suitable environments for conservation (true positive predictions) and restoration (false positive predictions). When the actual distribution of a threatened species or biotope is not known, it is much more costly for conservation to overlook a potential occurrence (low number of false positive predictions; Jiménez-Valverde & Lobo, 2007; Loiselle et al., 2003). However, as we know the actual distribution of the biotopes, we are interested in identifying sites that provide highly suitable environmental conditions for successful conservation or restoration. By using a lower or higher threshold to translate suitability values to a potential occurrence, conservation planning can adapt the sensitivity to environmental conditions and include raster cells with lower suitability for more options (see Figures S2a, S2b, S3a and S3b in Appendix S2) or select only raster cells with very suitable conditions (see Figures S2d and S3d in Appendix S2). We are convinced that the 0.5 threshold used in our study provides balanced results.

4.2 | Abiotic environmental variables driving the distribution of biotopes

For more than half of the biotopes, climate variables were the most important variables for the prediction of their distribution. This underlines the findings of other studies for different forest biotopes: Brzeziecki et al. (1995) showed that when the mean temperature was increased by more than 2°C, the entire distribution of biotopes on a national scale would change, and John et al. (2020) showed that rainfall was one of the most important variables in predicting the distribution of biotopes. Similarly, we found that *precipitation seasonality* and the *mean temperature of the wettest quarter* were important for a range of biotopes. The most likely explanation for the effect of climate variables on the distribution of biotopes is that the occurrence of the species characterising a biotope are driven by climate. In particular, the germination and growth of plants are highly dependent on climate (Kadereit et al., 2014). Further, climate affects plant growth indirectly by changing the availability of nutrients in the soil (Marschner & Rengel, 2012).

Despite the dominance of climatic variables, the distribution of some biotopes was also strongly influenced by several soil variables. Previous studies showed that also soil variables could determine the distribution of biotopes, either together with climate variables (Fischer et al., 2019) or by themselves. Our study found that *potassium* and *nitrogen* were important for "Large sedge meadows outside the siltation zone" and "Moist and wet tall herbaceous vegetation". Since these were biotopes with sufficient water availability and hence a lower importance of precipitation conditions, *potassium* and *nitrogen* became the limiting factor. High values of *coarse fragments*, i.e. the occurrence of rocks and gravel, characterise mainly the alpine regions and also its most dominant biotope, the "Alpine lawn" (Figure S1 in Appendix S2). As other alpine biotopes, however, were more restricted to specific areas in the alpine region (Figure S1 in Appendix S2), the model considered other variables as more important in these cases.

As shown by Franklin et al. (2013), the variable's importance may change due to the selected spatial resolution; however, even if some models considered soil variables more important than climate variables, we have to take into account that this is only possible if climate conditions are suitable. Only under suitable climate conditions may other environmental conditions become limiting in the distribution of a biotope (see, e.g. P- and N-limitation in terrestrial ecosystems; Vitousek et al. (2010)). This indicates, similar to other studies (e.g. Beaumont et al., 2011; Brzeziecki et al., 1995; Hickler et al., 2012; John et al., 2020), that especially future climate change may have a strong effect on these biotopes, which will probably force them to shift to more environmentally suitable areas (Hof et al., 2011).

4.3 | How land use drives the observed distribution of biotopes

In addition to the abiotic environmental conditions, which describe the potential of a biotope to occur in an area, anthropogenic land use and disturbance often influence the realised occurrence (Franklin, 1995). Consequently, the observed distribution of biotopes is the product of suitable abiotic conditions for a biotope and the outcome of the applied anthropogenic land-use or disturbance regimes, which in turn depend on human decisions that are often guided by economic factors. However, anthropogenic effects may not only establish a biotope in a suitable area but can also displace a biotope even if suitable abiotic conditions are given. Indication for this displacement is also evident in the results of our study. For all biotopes, we predicted a broader potential distribution based on environmental conditions compared with the observed distribution (Figure S1 in Appendix S2). This indicates that the potential of the biotopes based on abiotic conditions is not fully realised, most likely because of displacement effects of anthropogenic land use or disturbance. Instead, we see the realisation of other land-use types (e.g. agriculture with 46.3% of the Bavarian area; Bayrisches Landesamt für Statistik, 2020), which reduced the area for observed biotopes to just 5% of Bavaria (Rubanschi et al., 2022).

4.4 | Deviation between the observed and potential biotope distribution

Generally, most models had a high accuracy in predicting the distribution of biotopes. In particular, alpine and some forest biotopes, which were restricted to certain regions with unique environmental conditions, had the highest accuracy (Table 3). This high accuracy agrees with studies for other biotopes that are restricted to specific regions with characteristic environmental conditions (e.g. mangroves; John et al., 2020) or range-limited species (Hernandez et al., 2006).

For some modelled biotopes, the predictive accuracy was low (Table 3). This may happen when the model estimates a broad environmental niche for a biotope and predicts, therefore, a wider potential distribution in comparison to the observed distribution. One possible mechanism leading to a broad biotope environmental niche is when the biotope was defined by a variable species composition in the biotope classification key. An example is here the "Hedges, near natural" where the biotope classification key says that predominately "[…] native woody plant composition [were] mapped" (Lang & Zintl, 2018). Since different species compositions were accepted as "Hedges, near natural", the environmental niche of the biotope is the sum of multiple varying species niches, which might create a broad biotope niche (see Table S7 in Appendix S2). Another possible mechanism is that a biotope possibly occurred in many regions with varying environmental conditions (e.g. "Nutrient-poor old grasslands and fallow grasslands" with 89,116 observations). Consequently, the environmental niche of these frequent biotopes is broad, which leads to frequent predictions in biotope compositions (see Aune-Lundberg & Strand, 2017).

A second explanation for the lower model accuracy of some biotopes can be that additional variables affect the distribution of the biotope, which were not included in the models. Including these variables would probably have helped restricting the predictions more closely towards the observed biotope distribution. One such variable may be *water dynamics*, which could have improved the predictive accuracy for biotopes like "Watercourse companion trees, linear" by restricting its potential distribution to areas near water. Another such variable could be land use (see above). However, not only the direct displacement of a biotope is possible but also differences in land-use intensity, for example in "Species-rich extensive grassland". For this biotope, an increase in land-use intensity would change the species composition, potentially beyond the combinations defined as characteristic for the biotope and therefore changes the biotopes' actual distribution compared with its potential distribution. All these mechanisms are not exclusive and can act in combination when, for example, the definition of a biotope is not explicit, and the biotope also occurred in many regions with varying environmental conditions (e.g. "Hedges, near natural" with 197,402 observations).

While large deviations between the predicted and the observed distribution are generally interpreted as "poor" predictions (Pearce & Ferrier, 2000; Swets, 1988), we expected such deviations because we aimed at modelling the potential distribution of biotopes and not their actual distribution. This approach is based on our assumption that abiotic variables describe the potential distribution of a biotope, and other variables like land use and disturbance realise this potential distribution, which is a nested set within the potential distribution (Franklin, 1995). Therefore, we assumed that by leaving out land use and disturbance, we do not restrict our predictions to the actual distribution but rather picture the abiotic potential distribution, thus accepting lower accuracy values.

4.5 | The potential biotope composition

With the application of the 0.5 threshold, the models predicted large parts of Bavaria to be suitable for multiple biotopes (Figures 2 and 4), with little difference in suitability values between the most suitable biotopes (Table S2 in Appendix S3). Compared with other studies (e.g. Fischer et al., 2019; Zhou et al., 2016) we did not consider just the most suitable biotope but rather all potentially suitable biotopes for a raster cell. Given that raster cells were larger in area than the biotopes (average raster cell size 56.6 ha vs. biotope polygon size 0.51 ha; Rubanschi et al., 2022) and that the environmental values represent a mean value for a raster cell ignoring heterogeneity or fluctuations within the raster cell, we assume that our results provide an extended picture beyond the realised biotope distribution compared with previous studies. Especially, the potential co-occurrence of biotopes due to heterogeneity of environmental conditions within the large raster cells was nicely shown by Seo et al. (2009). There, they modelled the suitability for the same species at different spatial resolutions showing fluctuation in suitability within large raster cells. Thus, different locations within a raster cell may be suitable for different biotopes. Based on this, one could assume that with a sufficiently high resolution of the environmental variables only one biotope may be projected per raster cell. However, even then, it is important to consider for each raster cell multiple biotopes suitable because external drivers like land use may prefer one suitable biotope over the others.

When we further analyse the potential biotope composition, we see that biotopes with broad environmental niches were found in all potential biotope compositions with high proportion independent of the number of suitable biotopes in the composition (Figure 3). However, they were often displaced by biotopes with narrower niches if the raster cells were suitable for just a small number of biotopes. In these raster cells, the environmental conditions were either so unique that just the biotopes with a narrow niche could occur, or the environmental conditions were just suitable for biotopes with broad environmental niches.

In contrast to the regions that were potentially suitable for several different biotopes, based on our selected threshold, 8% of the Bavarian area was considered unsuitable for any of the biotopes (Figures 2 and 4). In these regions, the Bavarian biotope mapping dataset generally observed no or only a small number of the biotopes (cf., Rubanschi et al., 2022), probably due to displacement by agriculture or forestry (cf., Agency European Environment, 2020). If the biotopes were displaced, these regions may be suitable for one of the biotopes, but the models cannot assign the occurring environmental condition to any biotope and therefore consider these areas unsuitable. However, even when these areas would be unsuitable for any of the 29 analysed biotopes, it does not indicate that no biotope can occur there. Some of the excluded rare and/or aquatic biotopes could occur there as strongly differing environmental conditions might characterise these biotopes compared with the set of biotopes modelled here.

Since we used Bavarian biotopes and the Federal State of Bavarian as a study area, together with current environmental conditions, we created reasonable models within the study's boundaries (Fitzpatrick & Hargrove, 2009). To use the models outside of Bavaria and ensure reasonable predictions, the new environmental conditions would need to be in the models' predictability (Fitzpatrick & Hargrove, 2009). Thus, it is challenging to scale up the predictions of Bavarian biotopes for comparison with other potential distribution studies on a continental scale (e.g. Jiménez-Alfaro et al., 2018). Further, even though the biotopes of those studies may share similarities with Bavarian biotopes, they may not reflect the specific characteristics of the locally adapted Bavarian biotopes since they operate on a greater scale. While studies on such a great scale (e.g. Jiménez-Alfaro et al., 2018) can provide a general overview, local conservation and restoration planning need information about locally adapted biotopes and their potential distribution that a study like ours can provide.

4.6 | Implications for biodiversity assessments, conservation and restoration

Our results emphasised that the potential biotope richness of a raster cell could be even higher than the observed. Since this potential richness is based on a selection of 29 biotopes, these values should be interpreted as lower bounds of potential richness and may be even higher when taking into account the rare and/or aquatic biotopes that were not included in this study. With this new insight into the potential biotope composition, it will be possible to better assess the potential biodiversity since the biotopes are proxies for characteristic species compositions. Such assessments become possible when considering not only the biotope with the highest suitability in a raster cell but rather a composition of similarly suitable biotopes.

The maps we provide can support conservation planning by identifying areas where the protection of specific biotopes may be promising due to the high suitability of the environmental conditions. Extending conservation planning from the biotope with the highest suitability in an area towards a composition of similarly suitable biotopes enables to cover potential uncertainties due to fluctuation in land-use and disturbance regimes. Especially, regions that potentially host many different biotopes could be considered as priority areas for conservation. The potential biotope composition could even be further evaluated by using different climate scenarios to identify which biotopes or regions may be resilient to oncoming changes in climate.

From the restoration point of view, our results can be used to find areas suitable to reintroduce biotopes or whole biotope compositions. Restoration should be prioritised especially in areas where the number of potentially occurring biotopes was much larger than the observed biotope number. With this, we would increase the biotope diversity in an area and ensure the long-term persistence of the biotopes by increasing their entire distribution. Further, our models enable us to quantitatively include climate change into restoration planning. Applying different climate scenarios, the models can be used to identify regions where environmental conditions may become suitable and support the shifts of biotopes to more suitable environmental conditions by introducing biotopes.

In conclusion, our study showed that considering not only the most suitable biotope but rather a composition of potentially suitable biotopes in an area could provide important information for adapted conservation and restoration planning. Future research regarding biotope compositions is comparing the potential and realised biotope compositions and forecasting the impacts of future climate change on potential biotope distributions and composition.

ACKNOWLEDGEMENTS

This study was funded by the Bavarian Ministry of Science and the Arts in the context of the Bavarian Climate Research Network (bayklif) via projects BLIZ (S.R., S.T.M. and W.W.W.) and mintbio (C.H.) and supported by Deutsche Forschungsgemeinschaft (DFG) through TUM International Graduate School of Science and Engineering (IGSSE), GSC 81. We thank the Bavarian Environment Agency (Bayerisches Landesamt für Umwelt) for the providence of the data. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

None.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available: Bavarian biotope mapping https://www.lfu.bayern.de/umweltdaten/geodatendienste/pretty_downloaddienst.htm?dld=biotopkartierung accessed 31.03.2020; soil chemical properties at https://esdac.jrc.ec.europa.eu/content/chemical-properties-european-scale-based-lucas-topsoil-data accessed 02.02.2020; soil physical properties at https://esdac.jrc.ec.europa.eu/content/topsoil-physical-properties-europe-based-lucas-topsoil-data accessed 02.02.2020; climate variables https://www.worldclim.org/data/worldclim21.html accessed 23.05.2020.

ORCID

Sven Rubanschi https://orcid.org/0000-0002-8301-2730
Sebastian T. Meyer https://orcid.org/0000-0003-0833-1472
Christian Hof https://orcid.org/0000-0002-7763-1885
Wolfgang W. Weisser https://orcid.org/0000-0002-2757-8959

REFERENCES

Agency European Environment. (2020). *Corine Land Cover (CLC) 2018, Version 2020_20u1*.

Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, *43*(6), 1223–1232. https://doi.org/10.1111/j.1365-2664.2006.01214.x

Aune-Lundberg, L., & Strand, G.-H. (2017). Composition and spatial structure of dwarf shrub heath in Norway. *Norsk Geografisk Tidsskrift – Norwegian Journal of Geography*, *71*(1), 1–11. https://doi.org/10.1080/00291951.2017.1291536

Bailey, S.-A., Haines-Young, R., & Watkins, C. (2002). Species presence in fragmented landscapes: Modelling of species requirements at the national level. *Biological Conservation*, *108*(3), 307–316. https://doi.org/10.1016/S0006-3207(02)00119-2

Ballabio, C., Lugato, E., Fernández-Ugalde, O., Orgiazzi, A., Jones, A., Borrelli, P., Montanarella, L., & Panagos, P. (2019). Mapping LUCAS topsoil chemical properties at European scale using Gaussian process regression. *Geoderma*, *355*, 113912. https://doi.org/10.1016/j.geoderma.2019.113912

Ballabio, C., Panagos, P., & Monatanarella, L. (2016). Mapping topsoil physical properties at European scale using the LUCAS database. *Geoderma*, *261*, 110–123. https://doi.org/10.1016/j.geoderma.2015.07.006

Bayrisches Landesamt für Statistik. (2020). *Flächenerhebung nach Art der tatsächlichen Nutzung in Bayern zum Stichtag 31. Dezember 2016*. Bayrisches Landesamt für Statistik.

Beaumont, L. J., Pitman, A., Perkins, S., Zimmermann, N. E., Yoccoz, N. G., & Thuiller, W. (2011). Impacts of climate change on the world's most exceptional ecoregions. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(6), 2306–2311. https://doi.org/10.1073/pnas.1007217108

Brzeziecki, B., Kienast, F., & Wildi, O. (1995). Modelling potential impacts of climate change on the spatial distribution of zonal forest communities in Switzerland. *Journal of Vegetation Science*, *6*(2), 257–268. https://doi.org/10.2307/3236221

Chytrý, M., Tichý, L., Hennekens, S. M., Knollová, I., Janssen, J. A. M., Rodwell, J. S., Peterka, T., Marcenò, C., Landucci, F., Danihelka, J., Hájek, M., Dengler, J., Novák, P., Zukal, D., Jiménez-Alfaro, B., Mucina, L., Abdulhak, S., Aćić, S., Agrillo, E., … Schaminée, J. H. J. (2020). EUNIS habitat classification: Expert system, characteristic species combinations and distribution maps of European habitats. *Applied Vegetation Science*, *23*(4), 648–675. https://doi.org/10.1111/avsc.12519

Colwell, R. K., & Rangel, T. F. (2009). Hutchinson's duality: The once and future niche. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(Suppl 2), 19651–19658. https://doi.org/10.1073/pnas.0901650106

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Efron, B. (1992). Bootstrap methods: Another look at the jackknife. In S. Kotz & N. L. Johnson (Eds.), *Breakthroughs in statistics: Methodology and distribution* (pp. 569–593). Springer. https://doi.org/10.1007/978-1-4612-4380-9_41

Elith, J., Graham, C., Valavi, R., Abegg, M., Bruce, C., Ford, A., Guisan, A., Hijmans, R. J., Huettmann, F., Lohmann, L., Loiselle, B., Moritz, C., Overton, J., Peterson, A. T., Phillips, S., Richardson, K., Williams, S., Wiser, S. K., Wohlgemuth, T., & Zimmermann, N. E. (2020). Presence-only and presence-absence data for comparing species distribution modeling methods. *Biodiversity Informatics*, *15*(2), 69–80. https://doi.org/10.17161/bi.v15i2.13384

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (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

European Community. (1992). Council Directive 92/43/EEC of 21 may 1992 on the conservation of natural habitats and of wild fauna and flora (OJ L 206 22.07.1992 p. 7). In P. Sands & P. Galizzi (Eds.), *Documents in European community environmental law* (pp. 568–583). Cambridge University Press. https://doi.org/10.1017/CBO9780511610851.039

European Environment Agency. (2016). *European Digital Elevation Model (EU-DEM), version 1.1*.

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Fischer, H. S., Michler, B., & Fischer, A. (2019). High resolution predictive modelling of potential natural vegetation under recent site conditions and future climate scenarios: Case study Bavaria. *Tuexenia*, *39*, 9–40. https://doi.org/10.14471/2018.39.001

Fitzpatrick, M. C., & Hargrove, W. W. (2009). The projection of species distribution models and the problem of non-analog climate. *Biodiversity and Conservation*, *18*(8), 2255–2261. https://doi.org/10.1007/s10531-009-9584-8

Franklin, J. (1995). Predictive vegetation mapping: Geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography: Earth and Environment*, *19*(4), 474–499. https://doi.org/10.1177/030913339501900403

Franklin, J., Davis, F. W., Ikegami, M., Syphard, A. D., Flint, L. E., Flint, A. L., & Hannah, L. (2013). Modeling plant species distributions under future climates: How fine scale do climate projections need to be? *Global Change Biology*, *19*(2), 473–483. https://doi.org/10.1111/gcb.12051

Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). *Habitat suitability and distribution models*. Cambridge University Press. https://doi.org/10.1017/9781139028271

Halvorsen, R. (2012). A gradient analytic perspective on distribution modelling. *Sommerfeltia*, *35*(1), 1–165. https://doi.org/10.2478/v10208-011-0015-3

Hemsing, L. Ø., & Bryn, A. (2012). Three methods for modelling potential natural vegetation (PNV) compared: A methodological case study from south-central Norway. *Norsk Geografisk Tidsskrift – Norwegian Journal of Geography*, *66*(1), 11–29. https://doi.org/10.1080/00291951.2011.644321

Hernandez, P. A., Graham, C. H., Master, L. L., & Albert, D. L. (2006). The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, *29*(5), 773–785. https://doi.org/10.1111/j.0906-7590.2006.04700.x

Hickler, T., Vohland, K., Feehan, J., Miller, P. A., Smith, B., Costa, L., Giesecke, T., Fronzek, S., Carter, T. R., Cramer, W., Kühn, I., & Sykes, M. T. (2012). Projecting the future distribution of European potential natural vegetation zones with a generalized, tree species-based dynamic vegetation model. *Global Ecology and Biogeography*, *21*(1), 50–63. https://doi.org/10.1111/j.1466-8238.2010.00613.x

Hijmans, R. J., Phillips, S. J., Leathwick, J., & Elith, J. (2017). *Dismo: Species distribution modeling (R package version 1.1-4)*. https://cran.r-project.org/package=dismo

Hirzel, A. H., Le Lay, G., Helfer, V., Randin, C., & Guisan, A. (2006). Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, *199*(2), 142–152. https://doi.org/10.1016/j.ecolmodel.2006.05.017

Hof, C. (2021). Towards more integration of physiology, dispersal and land-use change to understand the responses of species to climate change. *Journal of Experimental Biology*, *224*(Suppl 1), 1–14. https://doi.org/10.1242/jeb.238352

Hof, C., Levinsky, I., Araujo, M. B., & Rahbek, C. (2011). Rethinking species' ability to cope with rapid climate change. *Global Change Biology*, *17*(9), 2987–2990. https://doi.org/10.1111/j.1365-2486.2011.02418.x

IPBES. (2019). *Global assessment report on biodiversity and ecosystem services of the intergovernmental science-policy platform on biodiversity and ecosystem services* (E. S. Brondizio, J. Settele, S. Diaz, & H. T. Ngo (Eds.)). IPBES Secretariat. https://doi.org/10.5281/ZENODO.5517154

IUCN. (2021). *The IUCN red list of threatened species. Version 2021-2*. https://www.iucnredlist.org

Jiménez-Alfaro, B., Suárez-Seoane, S., Chytrý, M., Hennekens, S. M., Willner, W., Hájek, M., Agrillo, E., Álvarez-Martínez, J. M., Bergamini, A., Brisse, H., Brunet, J., Casella, L., Dítě, D., Font, X., Gillet, F., Hájková, P., Jansen, F., Jandt, U., Kącki, Z., … Tsiripidis, I. (2018). Modelling the distribution and compositional variation of plant communities at the continental scale. *Diversity and Distributions*, *24*(7), 978–990. https://doi.org/10.1111/ddi.12736

Jiménez-Valverde, A., & Lobo, J. M. (2007). Threshold criteria for conversion of probability of species presence to either–or presence–absence. *Acta Oecologica*, *31*(3), 361–369. https://doi.org/10.1016/j.actao.2007.02.001

John, E., Bunting, P., Hardy, A., Roberts, O., Giliba, R., & Silayo, D. S. (2020). Modelling the impact of climate change on Tanzanian forests. *Diversity and Distributions*, *26*(12), 1663–1686. https://doi.org/10.1111/ddi.13152

Kadereit, J. W., Körner, C., Kost, B., & Sonnewald, U. (2014). *Strasburger – Lehrbuch der Pflanzenwissenschaften* (37th ed.). Springer. https://doi.org/10.1007/978-3-642-54435-4

Keith, D. A., Elith, J., & Simpson, C. C. (2014). Predicting distribution changes of a mire ecosystem under future climates. *Diversity and Distributions*, *20*(4), 440–454. https://doi.org/10.1111/ddi.12173

Lang, A., & Zintl, R. (2018). *Kartieranleitung Biotopkartierung Bayern (inkl. Kartierung der Offenland-Lebensraumtypen der Fauna-Flora-Habitat-Richtlinie) Teil 2 – Biotoptyp*. UmweltSpezial.

Liu, C., Berry, P. M., Dawson, T. P., & Pearson, R. G. (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, *28*(3), 385–393. https://doi.org/10.1111/j.0906-7590.2005.03957.x

Liu, C., White, M., & Newell, G. (2011). Measuring and comparing the accuracy of species distribution models with presence-absence data. *Ecography*, *34*(2), 232–243. https://doi.org/10.1111/j.1600-0587.2010.06354.x

Lobo, J. M., Jiménez-Valverde, A., & Hortal, J. (2010). The uncertain nature of absences and their importance in species distribution modelling. *Ecography*, *33*(1), 103–114. https://doi.org/10.1111/j.1600-0587.2009.06039.x

Loidi, J., & Fernández-González, F. (2012). Potential natural vegetation: Reburying or reboring? *Journal of Vegetation Science*, *23*(3), 596–604. https://doi.org/10.1111/j.1654-1103.2012.01387.x

Loiselle, B. A., Howell, C. A., Graham, C., Goerck, J. M., Brooks, T., Smith, K. G., & Williams, P. H. (2003). Avoiding pitfalls of using species distribution models in conservation planning. *Conservation Biology*, *17*(6), 1591–1600. https://doi.org/10.1111/j.1523-1739.2003.00233.x

Marage, D., & Gégout, J.-C. (2009). Importance of soil nutrients in the distribution of forest communities on a large geographical scale. *Global Ecology and Biogeography*, *18*(1), 88–97. https://doi.org/10.1111/j.1466-8238.2008.00428.x

Marschner, P., & Rengel, Z. (2012). Nutrient availability in soils. In P. Marschner (Ed.), *Marschner's mineral nutrition of higher plants* (2nd ed., pp. 315–330). Academic Press. https://doi.org/10.1016/B978-0-12-384905-2.00012-1

Mason, S. J., & Graham, N. E. (2002). Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. *Quarterly Journal of the Royal Meteorological Society*, *128*(584), 2145–2166. https://doi.org/10.1256/003590002320603584

Matthies, D., Bräuer, I., Maibom, W., & Tscharntke, T. (2004). Population size and the risk of local extinction: Empirical evidence from rare plants. *Oikos*, *105*(3), 481–488. https://doi.org/10.1111/j.0030-1299.2004.12800.x

Maxwell, S. L., Fuller, R. A., Brooks, T. M., & Watson, J. E. M. (2016). Biodiversity: The ravages of guns, nets and bulldozers. *Nature*, *536*, 143–145. https://doi.org/10.1038/536143a

Merow, C., Smith, M. J., & Silander, J. A. (2013). A practical guide to MaxEnt for modeling species' distributions: What it does, and why inputs and settings matter. *Ecography*, *36*(10), 1058–1069. https://doi.org/10.1111/j.1600-0587.2013.07872.x

Naimi, B., Hamm, N. A. S., Groen, T. A., Skidmore, A. K., & Toxopeus, A. G. (2014). Where is positional uncertainty a problem for species distribution modelling? *Ecography*, *37*(2), 191–203. https://doi.org/10.1111/j.1600-0587.2013.00205.x

Newbold, T., Oppenheimer, P., Etard, A., & Williams, J. J. (2020). Tropical and Mediterranean biodiversity is disproportionately sensitive to land-use and climate change. *Nature Ecology & Evolution*, *4*, 1630–1638. https://doi.org/10.1038/s41559-020-01303-0

Pearce, J., & Ferrier, S. (2000). Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, *133*(3), 225–245. https://doi.org/10.1016/S0304-3800(00)00322-7

Phillips, S. J. (2017). *A brief tutorial on Maxent*. http://biodiversityinformatics.amnh.org/open_source/maxent/

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*(3–4), 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

QGIS Development Team. (2020). *QGIS geographic information system – Version 3.16.4 (3.16.4)*.

R Core Team. (2020). *R: A language and environment for statistical computing*.

Rubanschi, S., Meyer, S. T., & Weisser, W. W. (2022). Die bayerische Biotopkartierung - Eine unterschätzte Ressource zur Untersuchung der Auswirkungen des Klima- und Landnutzungswandels auf den Naturschutz. *Mitteilungen Der Fränkischen Geographischen Gesellschaft*, *67*, 11–18.

Santika, T. (2011). Assessing the effect of prevalence on the predictive performance of species distribution models using simulated data. *Global Ecology and Biogeography*, *20*(1), 181–192. https://doi.org/10.1111/j.1466-8238.2010.00581.x

Seo, C., Thorne, J. H., Hannah, L., & Thuiller, W. (2009). Scale effects in species distribution models: Implications for conservation planning under climate change. *Biology Letters*, *5*(1), 39–43. https://doi.org/10.1098/rsbl.2008.0476

Sillero, N., Ribeiro-Silva, J., & Arenas-Castro, S. (2022). Shifts in climatic realised niches of Iberian species. *Oikos*, *2022*(4), e08505. https://doi.org/10.1111/oik.08505

Stockwell, D. R. B., & Peterson, A. T. (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

Swets, J. (1988). Measuring the accuracy of diagnostic systems. *Science*, *240*(4857), 1285–1293. https://doi.org/10.1126/science.3287615

Tarkesh, M., & Jetschke, G. (2012). Comparison of six correlative models in predictive vegetation mapping on a local scale. *Environmental and Ecological Statistics*, *19*, 437–457. https://doi.org/10.1007/s10651-012-0194-3

Tüxen, R. (1956). Die heutige potentielle natürliche Vegetation als Gegenstand der Vegetationskartierung. *Angewandte Pflanzensoziologie*, *13*, 4–42.

Vitousek, P. M., Porder, S., Houlton, B. Z., & Chadwick, O. A. (2010). Terrestrial phosphorus limitation: Mechanisms, implications, and nitrogen–phosphorus interactions. *Ecological Applications*, *20*(1), 5–15. https://doi.org/10.1890/08-0127.1

Vogiatzakis, I. N., & Griffiths, G. H. (2006). A GIS-based empirical model for vegetation prediction in Lefka Ori, Crete. *Plant Ecology*, *184*, 311–323. https://doi.org/10.1007/s11258-005-9075-2

Warren, D. L., Matzke, N. J., & Iglesias, T. L. (2020). Evaluating presence-only species distribution models with discrimination accuracy is uninformative for many applications. *Journal of Biogeography*, *47*(1), 167–180. https://doi.org/10.1111/jbi.13705

Watson, J. E. M., Dudley, N., Segan, D. B., & Hockings, M. (2014). The performance and potential of protected areas. *Nature*, *515*, 67–73. https://doi.org/10.1038/nature13947

Woolf, A., Nielsen, C. K., Weber, T., & Gibbs-Kieninger, T. J. (2002). Statewide modeling of bobcat, *Lynx rufus*, habitat in Illinois, USA. *Biological Conservation*, *104*(2), 191–198. https://doi.org/10.1016/S0006-3207(01)00164-1

Zhou, J., Lai, L., Guan, T., Cai, W., Gao, N., Zhang, X., Yang, D., Cong, Z., & Zheng, Y. (2016). Comparison modeling for alpine vegetation distribution in an arid area. *Environmental Monitoring and Assessment*, *188*(408), 408. https://doi.org/10.1007/s10661-016-5417-x

Zurell, D., Zimmermann, N. E., Gross, H., Baltensweiler, A., Sattler, T., & Wüest, R. O. (2020). Testing species assemblage predictions from stacked and joint species distribution models. *Journal of Biogeography*, *47*(1), 101–113. https://doi.org/10.1111/jbi.13608

Zuur, A. F., Ieno, E. N., & Elphick, C. S. (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

BIOSKETCH

Sven Rubanschi is interested in distribution models for biotopes at various spatial–temporal scales and the integration of those models in the applied conservation and restoration biogeography. Further, he is interested in the effect of climate and land-use change on the potential distribution of biotopes.

Author contributions: S.R. led the writing with significant input from S.T.M., C.H. and W.W.W.; S.R. performed all analyses with significant input from S.T.M. All authors have reviewed the manuscript.

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: Rubanschi, S., Meyer, S. T., Hof, C., & Weisser, W. W. (2023). Modelling potential biotope composition on a regional scale revealed that climate variables are stronger drivers than soil variables. *Diversity and Distributions*, *29*, 492–508. https://doi.org/10.1111/ddi.13675

APPENDIX 1

CLASSIFICATION DESCRIPTION OF BIOTOPES INTO THEIR BIOTOPE TYPES

In the first step, we merged the biotopes, consisting of multiple "Lebensraumtypen". The differences between them were as explained in the biotope mapping manual marginal (Lang & Zintl, 2018). With this step, we reduced the number of biotopes from 200 to 108. In the second step, we classified these biotopes into similar vegetational types based on the biotope mapping manual (Lang & Zintl, 2018). We distinguished between aquatic vegetation, bushes, forest, grassland, peatland, siltation zone, water body and "other" (Table S1 in Appendix S2).

We classified a biotope as aquatic vegetation when the biotope's species composition and the biotope mapping manual indicated that the vegetation is occurring within a water body. A biotope was classified as a bush biotope type if it was dominated by woody non-tree vegetation including shrub- and bushlands, hedges. If a biotope is dominated by tree species and shows a forest-typical vegetation structure, it was classified as a forest biotope type. Biotopes classified as grassland biotope types were dominated by grasses and herbs independent of the abiotic condition if there was no water accumulation for a longer period. If the biotope accumulated water for a longer period and produced peat, we classified the biotope as peatland independent of what higher vegetation was occurring within the biotope. The biotope type of siltation zone consisted of all biotopes directly adjacent to water bodies and relied on water dynamics. All mapped open water bodies, often with no vegetation, were classified as water body. Biotopes that did not fit in any of these biotope types because they were too diverse, only occurred at unique locations, were artificially created and dominated by anthropogenic actions or were no longer mapped and could not be assigned to another biotope were classified as "Other". The specific reason why a biotope was classified as "Other" can be found in Table S1 in Appendix S2 in the column "Classification explanation". Additionally, if the biotope mapping manual gave additional information about the biotope that affected our biotope type classification, we added the information also in the "Classification explanation" column of Table S1 in Appendix S2.

As described in the method section, we decided to concentrate in this study on terrestrial biotopes. Therefore, we excluded the aquatic vegetation, peatland, siltation zone, water body and "other" biotope types (Table S1 in Appendix S2). Additionally, we used a threshold of at least 500 observations per biotope. Which biotopes were excluded, and the reason for their exclusion was noted in the column "Reason of exclusion" in Table S1 in Appendix S2.