



Predicting species distributions based on incomplete survey data: the trade-off between precision and scale

Veronika Braunisch and Rudi Suchant

V. Braunisch (Veronika.Braunisch@forst.bwl.de), Forest Research Inst. of Baden-Wuerttemberg, Dept of Landscape Ecology, Wildlife Ecology Div., Wonnhaldestr. 4, DE-79100 Freiburg, Germany, and Univ. of Bern, Inst. of Ecology and Evolution, Conservation Biology, Baltzerstr. 6, CH-3012 Bern, Switzerland. – R. Suchant, Forest Research Inst. of Baden-Wuerttemberg, Dept of Landscape Ecology, Wildlife Ecology Div., Wonnhaldestr. 4, DE-79100 Freiburg, Germany.

Systematic species surveys over large areas are mostly not affordable, constraining conservation planners to make best use of incomplete data. Spatially explicit species distribution models (SDM) may be useful to detect and compensate for incomplete information. SDMs can either be based on standardized, systematic sampling in a restricted subarea, or – as a cost-effective alternative – on data haphazardly collated by “volunteer-based monitoring schemes” (VMS), area-wide but inherently biased and of heterogeneous spatial precision. Using data on capercaillie *Tetrao urogallus*, we evaluated the capacity of SDMs generated from incomplete survey data to localise unknown areas inhabited by the species and to predict relative local observation density. Addressing the trade-off between data precision, sample size and spatial extent of the sampling area, we compared three different sampling strategies: VMS-data collected throughout the whole study area (7000 km²) using either 1) exact locations or 2) locations aggregated to grid cells of the size of an average individual home range, and 3) systematic transect counts conducted within a small subarea (23.8 km²). For each strategy, we compared two sample sizes and two modelling methods (ENFA and Maxent), which were evaluated using cross-validation and independent data. Models based on VMS-data (strategies 1 and 2) performed equally well in predicting relative observation density and in localizing “unknown” occurrences. They always outperformed strategy 3-models, irrespective of sample size and modelling method, partly because the VMS-data provided the more comprehensive clues for setting the discrimination-threshold for predicting presence or absence. Accounting for potential errors due to extrapolation (e.g. projections outside the environmental domain or potentially biasing variables) reduced, but did not fully compensate for the observed discrepancies. As they cover a broader range of species-habitat relations, the area-wide data achieved a better model quality with less a-priori knowledge. Furthermore, in a highly mobile species like capercaillie a sampling resolution corresponding to an individuals’ home range can lead to equally good predictions as the use of exact locations. Consequently, when a trade-off between the sampling effort and the spatial extent of the sampling area is necessary, less precise data unsystematically collected over a large representative region are preferable to systematically sampled data from a restricted region.

Assessing the spatial distribution of rare and endangered species is essential for efficient conservation management (Margoluis and Salafsky 1998, Stem et al. 2005). To comply with the requirements of national and international biodiversity legislation and conservation programmes, a variety of survey schemes have been developed (Stem et al. 2005). These are rarely consistent in methodology and tend to vary with species’ attributes, environmental constraints and available financial resources. Vertebrate species can be particularly challenging to survey, owing to their high spatial requirements (with regard to both the individual home range and the area required by a viable population), elusiveness, spatio-temporal mobility and low detectability, with recurrent, large-scale surveys often necessary to appraise their distribution status (Kéry and Schmid 2004).

As systematic, repeated sampling over wide areas is often out of scope, most species surveys are restricted to small subregions. A cost-effective alternative may be the reliance on casual species records reported by knowledgeable volunteers, but volunteer-based monitoring schemes’ (VMS) have their caveats: not all of the areas inhabited by a species are recorded reliably, in particular information about the species’ absence is missing. Observation densities can be biased towards areas frequently visited by observers; and data quality is often very heterogeneous, particularly with regard to the spatial resolution.

To enhance the informational output of monitoring programmes for conservation purposes, the use of spatially explicit species distribution models (SDM) has been suggested (Brotons et al. 2007). Providing area-wide predictions about species presence, SDMs can substantially

increase the effectiveness of sampling strategies (Guisan et al. 2005, Singh et al. 2009) and enhance the detection of unknown populations of rare species (Engler et al. 2004), and thus offer the potential to compensate for biased or incomplete survey information.

A variety of methods exist to model species distributions (Guisan and Zimmermann 2000, Elith et al. 2006). Yet, as reliable information on species absence is often lacking, a focus has recently been placed on methods that rely on presence-only data (Hirzel et al. 2002, Pearce and Boyce 2006). Presence-only models can either be based on environmental envelopes characterising the environmental conditions at the species locations (Busby 1991, Carpenter et al. 1993) or on techniques that contrast these conditions with conditions prevailing at random locations (“pseudo-absences”, Guisan et al. 2005, Pearce and Boyce 2006, Zarnetske et al. 2007) or throughout the study area (Hirzel et al. 2002, Johnson et al. 2006, Lele and Keim 2006, Lele 2009).

The effect of input-data quality and quantity on model performance is addressed by several studies (Engler et al. 2004, Guisan et al. 2005, Edwards et al. 2006, Hernandez et al. 2006). Most of them concern data collected in systematic field surveys, others use haphazard VMS-data, but there is a lack of studies comparing both strategies. This may be due to the general assumption that models based on arbitrarily collected data perform worse than those based on well-designed, systematic sampling (Schreuder et al. 2001), which has been confirmed for even sample sizes drawn from the same sampling area (Edwards et al. 2006). In conservation practice, however, due to typical financial, time and manpower constraints, a trade-off is often necessary between the sampling precision, the sample size and the extent of the sampling area. Consequently, for a conservation manager it is crucial to know whether it is preferable to invest limited resources into systematic, precise, but local-scale sampling or to base models and conservation strategies on VMS data, area-wide but suffering from the described biases.

We address this question using capercaillie *Tetrao urogallus* data collected in the Black Forest, Germany. Capercaillie *Tetrao urogallus*, an indicator species of large, well structured boreal and montane forests, is highly endangered throughout its central European range (Storch 2000). This species has high spatial requirements, with individual annual home ranges of 100–1000 ha (Storch 1995) and a viable population requiring at least 15 000 ha (Suchant and Braunisch 2004). Accordingly, a VMS has been launched, based on a year-round collation of any signs of capercaillie presence over the whole ecoregion (Braunisch and Suchant 2006). In addition, for the purpose of the comparison, systematic transect counts have been conducted in a small subarea of the same ecoregion.

We compare SDMs based on these sampling strategies with regard to their capacity to localise unknown areas of species occurrence and to predict relative observation densities at a local scale. Addressing the trade-off between quality, quantity and spatial extent we pose the following questions: 1) which of the two sampling strategies provides the better outcome; VMS-data sampled across the whole study area or systematically sampled data from a small subarea? 2) How much greater is the informational value of

VMS-data with exact locations compared to the same data aggregated to a spatial resolution corresponding to an average individual home range size? 3) Do the results of 1) and 2) differ relative to the sample size, does one sampling strategy outperform the other in relation to the amount of occurrence data that can be expected? As the outcome of SDMs can further vary according to modelling procedure (Elith et al. 2006, Brotons et al. 2007), we compare two presence-only approaches, an ecological niche factor analysis (ENFA, Hirzel et al. 2002) and a maximum entropy model (Maxent, Phillips et al. 2006), both popular in applied conservation planning. By first simulating their application under “real conservation-management conditions” (i.e. absence of a-priori knowledge, decisions only based on the clues provided by input data and software), then comparing the outcome with the estimates from the approach with the highest predictive performance, we discuss the possibilities and pitfalls linked with processing incomplete survey data.

Methods

Study area

The study area encompasses the Black Forest in southwestern Germany, a forested mountain range of ca 7000 km², with an elevation ranging from 120 to 1493 m a.s.l. Capercaillie – once present in most of the forested parts of the study area – has undergone a strong population decline since the end of the 19th century. The current population of ca 600 individuals is distributed over a total area of 51 000 ha (Braunisch and Suchant 2006) arranged as over 100 fragmented patches due to topographic constraints and human land-use (Suchant and Braunisch 2004). For systematic sampling a subarea of 23.8 km² was delineated in the northwestern part of the study area (Fig. 1), with an elevation ranging from 400 to 1163 m a.s.l. The subarea was chosen so as to be as representative as possible with regard to both the environmental conditions given in the study area and used by the species (Fig. 2).

Capercaillie data

Capercaillie has been monitored in the Black Forest since 1988. All direct and indirect evidence of capercaillie presence provided by foresters, hunters, ornithologists, conservation volunteers and research personnel are collected. Every five years (1993, 1998, 2003) the species distribution is mapped based on all available data. We only use records where there is repeated evidence that a bird is using a patch. Specifically, patches are classified as “inhabited” when at least three capercaillie records occurring with a maximum distance of 1000 m from another have been collected within the preceding five year period, and are delineated by the minimum polygon encompassing these records (for details see: Braunisch and Suchant 2006). For this study, we randomly subdivided the inhabited patches of at least 100 ha into calibration and validation patches (Fig. 1), with the latter meant to simulate “unknown” areas of species presence. We used a total of

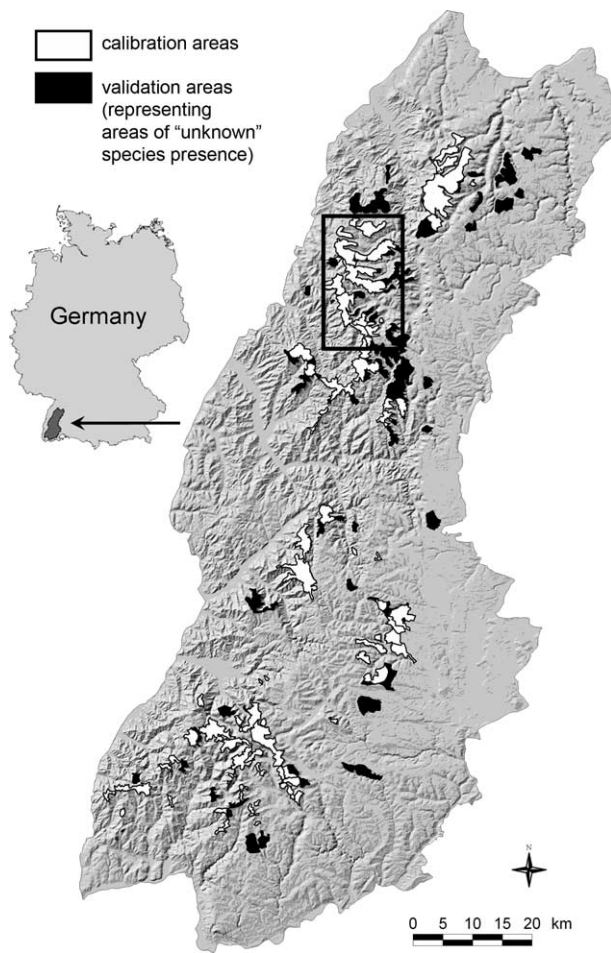


Figure 1. Distribution area of the capercaillie in the study area, randomly subdivided into calibration and validation areas (minimum size: 100 ha), with the latter used to simulate areas of “unknown” species presence. A minimum distance of 1 km² was applied between both categories. The black rectangle indicates the location of the subarea used for systematic transect counts.

1178 capercaillie records collected within these patches during the last monitoring period (1998–2003); n = 587 stemming from calibration patches and n = 600 from validation patches, with a minimum distance of 1 km between calibration and validation data. Data were selected randomly with a minimum distance of 300 m between all samples to reduce biases arising from spatial autocorrelation and from multiple, non-independent samples of the same individual (Thiel 2007), although this cannot be totally excluded as several individuals with overlapping home-ranges co-occur in a patch. Only the data pertaining to the calibration areas were used for model generation.

In addition, we conducted systematic transect counts in the northwestern subarea. Corresponding to the gradsect approach (Austin and Heyligers 1991), transects were established in forest stand units (i.e. homogeneously structured forest patches of a mean size of 3.5 ha) so that the total transect length crossing each forest stand was proportional to its size. Transect counts were conducted twice, in winter 2003/2004 and summer 2004, resulting in a total sample size of n = 1 108.

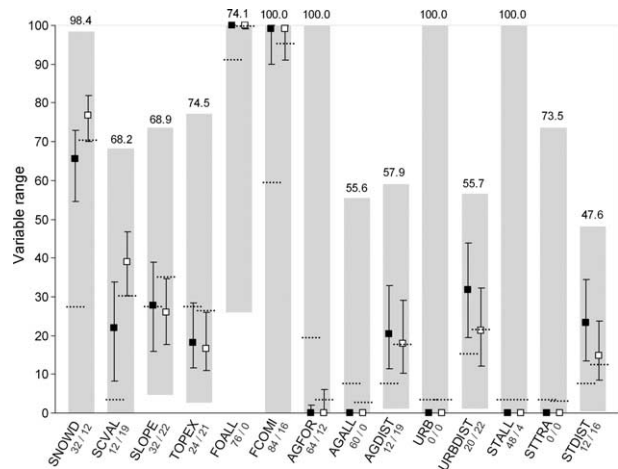


Figure 2. Environmental conditions and capercaillie distribution in the subarea compared to the study area. For each variable, the range of values given in the study area is set to 100. Grey bars illustrate the part of the variable range present in the subarea, with the respective proportions indicated above. In addition, the distribution of capercaillie samples (median and interquartile range) is shown. Black squares: VMS-data collected throughout the study area (dataset 1, n = 578), white squares: transect samples from the subarea (dataset 3, n = 1108), dotted line: median of variable in the respective sampling area. The interquartile ranges (study area/subarea) given with each variable.

We compared models based on three different sets of input data. The first (strategy 1) contained the VMS-data using the exact locations (n = 578). For the second (strategy 2) each of these samples was assigned to the next coordinate plane in the 1 × 1 km Gauss-Krüger graticule (Fig. 3), which was then treated as one single sample location independent of the number of samples occurring within the 1 km² grid cell. The 1 km² corresponds to the size of a small annual capercaillie home range (Storch 1995). This procedure reduced the original number of presence locations (n = 578) to n = 309. The third dataset (strategy 3) consisted of the data from the systematic transect counts in the subarea (n = 1108). To test whether the model results were sensitive to the sample size, we performed all analyses also with a reduced sample set of 60 presence points, randomly selected from each of the initial datasets.

Environmental variables

We included environmental variables (EV) (Table 1) that had been proven to explain capercaillie distribution in an earlier study (Braunisch and Suchant 2007). For each variable we prepared raster maps with a 30 × 30 m grid. To adjust the variable resolution to the resolution of the respective species data set (cf. Guisan and Thuiller 2005), we calculated the variable mean within circular moving windows of a size corresponding to the sampling resolution, that is, 9 ha (300 × 300 m) for the models based on the species data sets (1) and (3), and 1 km² for the model based on (2). As multinormality was required for the ENFA model, the variables were normalised using the Box-Cox algorithm (Box and Cox 1964). Maps were prepared in ArcView (ESRI 1996) and converted to IDRISI (Eastman

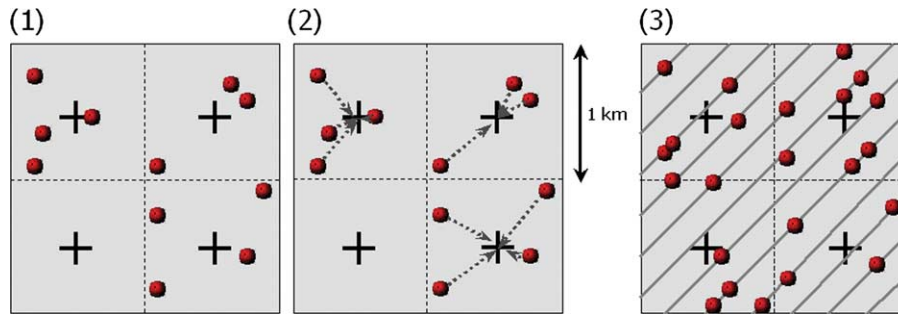


Figure 3. Schematic illustration of the three different sampling strategies compared in this study. (1) Data obtained by a volunteer-based monitoring scheme (VMS) across the whole study area (7000 km²), using the exact sample locations. (2) As (1) but with data assigned to the next coordinate plane (black cross) of the 1 km² Gauss-Krueger graticule. All samples within the respective 1 km² were aggregated to one single sample. (3) Systematic transect counts, conducted in the subarea (23,8 km²).

1990) for use in BIOMAPPER 4 (Hirzel et al. 2007) and into ASCII format for use in Maxent 3.2.1 (Phillips et al. 2006).

Statistical methods

We applied two modelling methods to each of the six data sets: an ecological niche factor analysis (ENFA, Hirzel et al. 2002) and a maximum entropy approach (Maxent, Phillips et al. 2006), both designed to cope with missing absence data. To simulate application under “conservation-reality” conditions (i.e. with no independent validation data available) models were first generated using only the information provided by the calibration data and the respective software. Then, in a second step, we compared

the models by evaluating their predictive power on the validation data.

Ecological niche factor analysis (ENFA)

The ecological niche factor analysis (ENFA) implemented in the software BIOMAPPER by Hirzel et al. (2002), is a profile based approach which compares the conditions prevailing on sites with proved species presence with those found across the entire study area. Comparable to a principal component analysis, all of the environmental variables are transformed to the same number of linearly uncorrelated factors. The first factor explains the species’ marginality, i.e. the difference between the average conditions on the sites where the species occurs (species distribution) and those across the entire study area (global

Table 1. Environmental variables included in the models.

Variable category	Variable description	Code	Unit	Data source	Original resolution
Landscape variables					
Climate	Number of days per year with snow cover >10 cm	SNOWD	days	DEM, modelled according to Schneider and Schönbein 2003	25 × 25 m
Soil conditions	Soil conditions, evaluated according to their potential to support suitable forest types	SCVAL	index (1–15)	Soil condition database	Vector data 1:10 000
Topography	Slope (Topex-to-distance index, distance: 2000 m. Low values indicate exposed sites)	SLOPE	degree	DEM	25 × 25 m
		TOPEX	(+/-) degree+1000	DEM, modelled according to Mitchell et al. 2001	25 × 25 m
Land-use variables					
Forest	Proportion of forest	FOALL	% of area	Landsat 5	30 × 30 m
	Proportion of coniferous and mixed forest	FCOMI	% of area	Landsat 5	30 × 30 m
	Proportion of forest-agriculture border area (200 m width)	AGFOR	% of area	Landsat 5	30 × 30 m
Agriculture	Proportion of agriculture	AGALL	% of area	Landsat 5	30 × 30 m
	Distance to agriculture	AGDIST	m	Landsat 5	30 × 30 m
Settlements	Proportion of settlements and urban area	URB	% of area	Landsat 5	30 × 30 m
	Distance to settlements and urban area	URBDIST	m	Landsat 5	30 × 30 m
Linear infrastructure	Proportion of area influenced by roads (plus 100 m buffer)	STALL	% of area	ATKIS	Vector data 1: 25 000
	Proportion of roads (plus 100 m buffer) weighted according to average traffic/day	STTRA	% of area × traffic-index	ATKIS/Ministry of Traffic	Vector data 1: 25 000
	Distance to roads	STDIST	m	ATKIS	Vector data 1: 25 000

distribution), indicating the position of the niche in the environmental space. The subsequent factors explain the specialisation, i.e. the ratio of global variance to species variance, indicating the niche breadth (Hirzel et al. 2002). The predicted frequency of species presence is calculated from the distribution of species data on the significant first factors, assigning to each cell in the study area a value, which is proportional to the distance of this cell to the species' optimum in the environmental factor space. The "area-adjusted median algorithm with extreme optimum" (Braunisch et al. 2008) was employed here, which was specifically developed for species in marginal habitats. The resulting scores range from 0 to 100, with 0 indicating conditions where the species is not present in the calibration data set and 100 representing conditions where the species is most frequently found. Assuming that these conditions represent the most suitable habitats, these scores are usually termed "habitat suitability (HS) scores", and so will be referred to in the following.

Maximum entropy model (Maxent)

Maxent is a machine-learning technique based on the principle of maximum entropy (Jaynes 1957) that was adapted for species distribution modelling by Phillips et al. (2004, 2006). The method seeks to find the probability distribution of species presence over all cells of the study area that best agrees with the constraints given by the observed species locations and, at same time, is closest possible to uniform (maximum entropy) in order to avoid any unfounded constraints. As predictors, the environmental variables and functions thereof (features) are used. We employed all available feature types namely linear, quadratic and product features, as well as threshold and hinge features (Phillips and Dudik 2008). As in logistic regression, each of these features is weighted by a constant. Starting with a uniform probability distribution, the weights are iteratively changed to converge to the probability distribution that best matches the distribution of the occurrence data set. Regularisation functions are employed, constraining the average value for a given feature so as to be close (i.e. within the confidence intervals) to its empirical average. As this is a generative approach, Maxent is especially advantageous when processing small and noisy data sets (Elith et al. 2006, Phillips et al. 2006). Maxent assigns a sample-likelihood to each grid cell within the study area. This probability distribution is exponential to the sum of the weighted features divided by a scaling constant, which ensures that the probability values range between 0 and 1 and sum up to 1. To predict the relative frequency of species presence, cumulative values were calculated, i.e. each cell was assigned the sum of the probabilities of this cell and all other grid cells in the study area with equal or lower probability multiplied by 100 (Phillips et al. 2006) to achieve values ranging between 0 and 100. Yet our results can be equally applied to the other output types of Maxent ("raw" or logistic output), as we only test the ordering of the predictions, which is consistent between output types. The models were run with a maximum of 500 iterations, a convergence threshold of 10^{-5} and a regularisation value of $\beta = 10^{-4}$.

Cross validation

The models' predictive power within their range of calibration was assessed by 10-fold cross validation (Fielding and Bell 1997) with pairwise differences between models tested by applying a Student's t-test on the 10 validation replicates. Two evaluation indices were calculated.

The area under the receiver operating characteristic (ROC) curve (AUC) is generally obtained by plotting sensitivity (i.e. the true positive rate, absence of omission error) against 1-specificity (i.e. the false positive rate, commission error) for all possible HS scores (Zweig and Campbell 1993, Fielding and Bell 1997). When no absence data are available, AUC-statistics can be calculated using a representative number of background cells instead (e.g. 10 000), randomly chosen from the study area (Wiley et al. 2003, Phillips et al. 2006). In this special case, the AUC measures the models' ability to discriminate between "presence" and "random" rather than between "presence" and "absence" and the maximum achievable AUC is always < 1 (Wiley et al. 2003). Consequently, as this maximum is unknown, AUC can only be employed for comparing models, not for assessing their absolute performance (Phillips et al. 2006).

The Continuous Boyce Index (CBI) (Boyce et al. 2002, Hirzel et al. 2006) was especially designed for presence-only models. It is obtained by plotting the relation between the predicted number of evaluation points (P) and the number of points expected by a random distribution (E) along the HS-value gradient (Fig. 3). The index, defined as the Spearman's R between P/E and HS, ranges from 0 (random) to 1 (perfect prediction), with negative values indicating models that predict worse than random.

Threshold selection

We used the results of the cross-validation to assess the best cut-off value for discriminating between predicted presence and absence. Either metric could be used to determine optimal thresholds, but we followed the defaults provided with the respective programmes. In the ENFA models, threshold-selection was based on the continuous Boyce curve (Supplementary material Fig. S1a). In adherence to Hirzel et al. (2006), we assigned "presence" to all HS values larger than the value at which the P/E, including its 90% confidence interval calculated from the 10 cross-validation replicates, was > 1 . This corresponds to the values for which the predicted frequency of species presence is significantly higher than expected by a random distribution.

In the Maxent models the threshold for discriminating between presence and absence was provided by the ROC curve: by shifting this threshold to larger HS values, the commission error will decrease but the omission error will increase (Fielding and Bell 1997) (Supplementary material Fig. S1b). Thus, when both error types are considered to be similarly problematic, the best cut-off value corresponds to the point on the ROC curve where sensitivity and specificity are maximised; that is, where the total amount of misclassification is minimised. We determined and averaged this value over the ten cross-validation replicates.

External evaluation

Predicting areas of “unknown” species presence

Cohen’s Kappa (κ) (Cohen 1960, Fielding and Bell 1997) was employed for testing the models’ capacity to localise the validation areas (representing “unknown” areas of species occurrence). In addition to the validation-data ($n = 600$) we randomly selected the same number of pseudo-absences (in the following referred to as “absence-data”) from the non-inhabited areas, applying a minimum distance of 1 km to all inhabited areas (“location”, Fig. 4a). The 1 km-distance was chosen based on a telemetry study showing that $>95\%$ of all locations of birds sporadically using areas outside of the mapped capercaillie areas remained within this distance (Braunisch and Thiel unpubl.). To test whether this buffer zone influenced the results, and to assess the model’s precision in predicting the capercaillie core areas, we created a second set of absence data ($n = 600$) randomly distributed within the 1 km buffer zone around the validation areas and with a minimum distance of 100 m from the presence points (“precision”, Fig. 4b). Kappa was calculated for all model variants. Pairwise differences between the κ -coefficients were assessed by first estimating the supposed “common” κ -value as described by Fleiss (1981), which was then used to test for equal κ -values on the Chi-square distribution. Kappa statistics were calculated using the ArcView 3.3 extension KAPPA STATS (Jenness and Wynne 2006).

Effect of threshold selection

We compared the models’ performance independent of an a priori chosen threshold using the validation presence-data with each of the two absence data sets. First, we calculated Cohen’s Kappa for all possible thresholds between 0 and 100, keeping the highest Kappa value (κ_{\max}) as a representation of the best possible model fit (Guisan et al. 1998). A comparison of κ_{\max} with the value obtained with the previously chosen threshold (κ_{thres}) showed the extent to which the model performance was influenced by threshold selection. Second, we calculated the AUC on both validation datasets using the R-package PresenceAbsence (Freeman and Moisen 2008).

Effect of extrapolation

Model extrapolations are prone to prediction errors, particularly when extrapolating beyond the environmental range used for calibration (Dormann 2007), or when variable importance differs between calibration and extrapolation area (Graf et al. 2006). To assess whether the relative performance of the locally calibrated models (strategy 3) were affected by such errors, we calculated two additional model variants for each sampling strategy: the first restricting predictions to the environmental domain of the subarea (64% of the study area), the second excluding the variable which differed most in importance between subarea and study area.

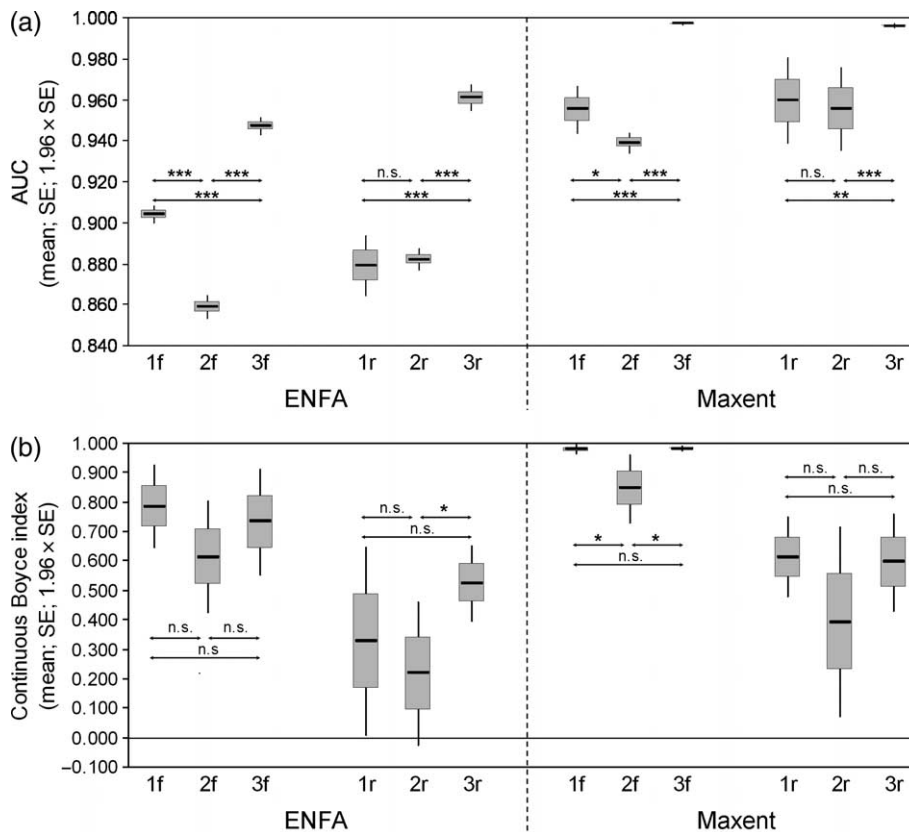


Figure 4. Results of the 10-fold cross-validation, comparing models based on data obtained with the three different sampling strategies (indicated by 1, 2, 3) using the full (f) and a reduced (r) data set. Models were evaluated using the area under the receiver operating characteristics curve AUC (a) and the Continuous Boyce index (b). Asterisks indicate the level of significance (*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$).

Predicting relative observation densities at a local scale

Finally, to evaluate the capability of the VMS-data models (strategy 1 and 2) to predict relative observation densities at a local scale within a forest stand unit, we compared the model predictions with the results of the systematic transect counts in the subarea. First, we calculated the mean HS score obtained with each of the models for each of the forest stand units ($n = 677$), which were then subdivided into ten groups according to their HS score. For each HS-interval we calculated the number of observations per unit area and related them to the modelled HS scores using Spearman's rank correlation. Pairwise differences between Spearman's Rs were then assessed using a two-sided test for differences between correlation coefficients (STATISTICA, StatSoft 2006).

Results

Relative importance of predictor variables

Although the predictor importance varied between models, the same variables always ranked amongst the most important (Supplementary material Table S1). Capercaillie presence was most positively correlated with cold climate conditions (SNOWD), a high proportion of forest (FOALL) and a low proportion of agricultural land (AGALL). Furthermore, areas with a high soil condition index (SCVAL) were preferred in all cases, but in models calibrated in the subarea (3), the relative importance of SCVAL always ranked much higher than in the models based on the samples from the whole study area (1, 2).

Cross-validation: model performance within the areas of calibration

Within their area of calibration, the models generated from systematically collected data (3) performed best, followed by those of sampling strategy 1 and 2. However, the clearness of this pattern differed between the two measured aspects of model quality: whereas AUC quantifies the models' performance to make binary predictions of presence and absence, integrated over all possible thresholds, CBI evaluates the models' capability to predict a smooth, consistent increase of observation frequencies over the HS-gradient (Hirzel et al. 2006). Strategy 3 models significantly ranked first regarding the former aspect (AUC, Fig. 4a), but there was no evidence of a difference to strategy-1 models regarding the latter (CBI, Fig. 4b). In addition, Maxent significantly outperformed the corresponding ENFA-models with regard to AUC, a similar trend was observed for CBI. Sample size negatively affected CBI, but not AUC, with no evidence of a difference between ENFA and Maxent found in this regard (Fig. 4a, b).

Predicting areas of "unknown" species presence

The spatial pattern and range of the predicted capercaillie distribution differed considerably between models (Fig. 5). Whereas the strategy 2 models tended to overestimate the

capercaillie distribution, fewer and smaller areas were predicted by the strategy 3 models, particularly in southern regions distant to the sampling area. This applied to both modelling approaches, but the pattern was more pronounced in Maxent. The VMS-data models using precise locations (1) and locations aggregated to a 1 km^2 grid (2) performed equally well in localising the validation areas ("location", Fig. 6a) and in predicting their spatial extent ("precision", Fig. 6b). According to Landis and Koch (1977), their accuracy was good ($0.4 > \kappa > 0.75$) to excellent ($\kappa > 0.75$) and they always outperformed the strategy 3 models, which generally revealed a poor agreement with the evaluation data ($\kappa < 0.4$). This finding was independent of the sample size and consistent over both methods (ENFA and Maxent). However, while ENFA and Maxent performed similarly on the VMS-data, Maxent models were always inferior to ENFA models when extrapolating from the locally and systematically sampled data (3). No advantage was recorded for either of the two approaches when processing the reduced data sets.

Influence of threshold selection

The threshold-independent evaluation generally supported the results described (Table 2) but the discrepancies between the VMS-data models (1, 2) and those extrapolated from local sampling (3) were less pronounced. Models of data set 1 and 2 performed excellently ($\text{AUC} > 0.9$, cf. Hosmer and Lemeshow 2000) in the first validation step ("location") and acceptably ($\text{AUC} > 0.7$) to good ($\text{AUC} > 0.8$) in the second ("precision"). Models of (3) were rated acceptable to good in the first step, and acceptable to poor ($0.6 < \text{AUC} < 0.7$) in the second evaluation step. In the models based on strategies (1) and (2), the accuracy obtained with the pre-selected threshold (κ_{thres}) was on average 80% (range: 52–99%) of the maximum achievable accuracy (κ_{max}), but it deviated $> 50\%$ from the optimum in strategy 3 models. This indicated that their inferiority was not solely related to a lower generalisation capacity, but also to a greater inaccuracy in providing transferable clues for threshold-selection.

Effect of extrapolation

Restricting the extrapolations to the environmental domain of the subarea reduced the qualitative differences between the strategy 3 models and the corresponding VMS-based models (Table 3). Omitting the variable "soil conditions" (SCVAL, Supplementary material Table S1) only reduced differences in AUC-values in the Maxent models. However, both modifications were associated with a decrease in predictive power in the corresponding VMS-models.

Predicting relative observation densities at a local scale

All models built from VMS-data were very successful in predicting relative observation densities at the local (forest stand) scale, with Spearman's correlation coefficients ranging between 0.745 and 0.988 (Table 4). Exact sample

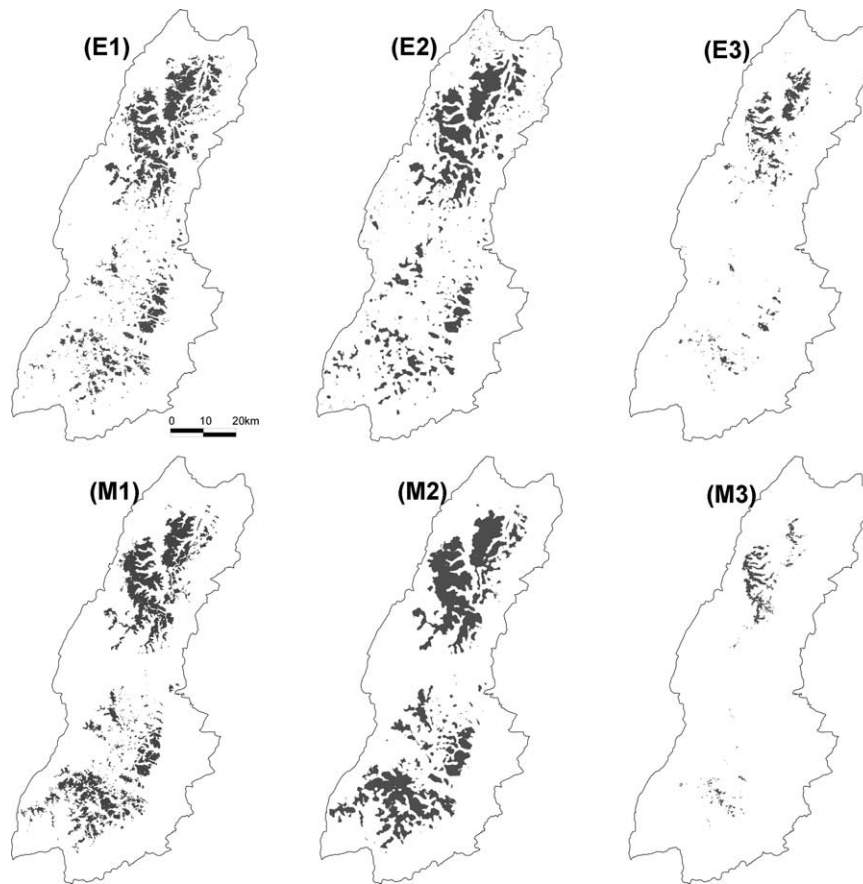


Figure 5. Species distribution maps predicted by ENFA (E) and Maxent (M) models based on species data obtained using three different sampling strategies. (1) Volunteer-based monitoring data with precise locations, (2) volunteer-based monitoring data aggregated to a 1 km² grid, and (3) systematically sampled data from the subarea.

locations tended to perform better than samples aggregated to a home range-sized grid, and Maxent generally provided slightly better results than ENFA, but these differences were not significant. Reducing the sample size to sixty presence points did not lead to a significant decrease in model predictions either.

Discussion

Assessing species distributions in conservation practice usually entails two major challenges: first, optimising sampling strategies with limited resources; second, maximizing information return from limited data. Both aspects were addressed here. We first tested the effect of spatially vs qualitatively limited sampling on species distribution models; second, we evaluated the performance of these models to fill the gaps inherent to fragmentary survey data (predicting unrecognized, but inhabited areas as well as relative observation densities). However, as the capercaillie areas we used to simulate areas with unknown species presence were also mainly inferred from VMS-data, they may be suspected to provide an insufficient, as potentially incomplete, basis for model evaluation. We consider this to be a minor problem here. First, the Black Forest is highly accessible and the monitoring programme is supported by a large number of well-trained people. Most parts of the study area are sampled by several people simultaneously

(Suchant et al. 2005). Therefore, we expect the probability of the species having remained undetected within entire inhabited patches to be very low. Second, systematic samplings (e.g. this study) and telemetry studies (Thiel et al. 2008, Braunisch and Thiel unpubl.) in different pilot areas of the Black Forest were found to match closely the distribution mapped from VMS-data, indicating a good spatial precision. Nevertheless, as we do not know the true species distribution with absolute certainty, we can only compare, but not test the models' accuracy.

Model generation: precision vs scale

Choosing an appropriate spatial scale is a central issue in SDM building (Wiens 2002). We compared sampling strategies that differed with regard to both relevant dimensions of scale, spatial resolution – or precision – (strategy 1 and 3 vs 2), and extent (strategy 1 and 2 vs 3).

1) Sampling resolution

Sampling resolution should optimally be chosen as to be coherent with the resolution of the predictor variables, and to correspond to the scale relevant for habitat selection processes (Guisan and Thuiller 2005, Guisan et al. 2007). Whereas for sessile organisms all necessary features have to be present at the same location, in highly mobile species like

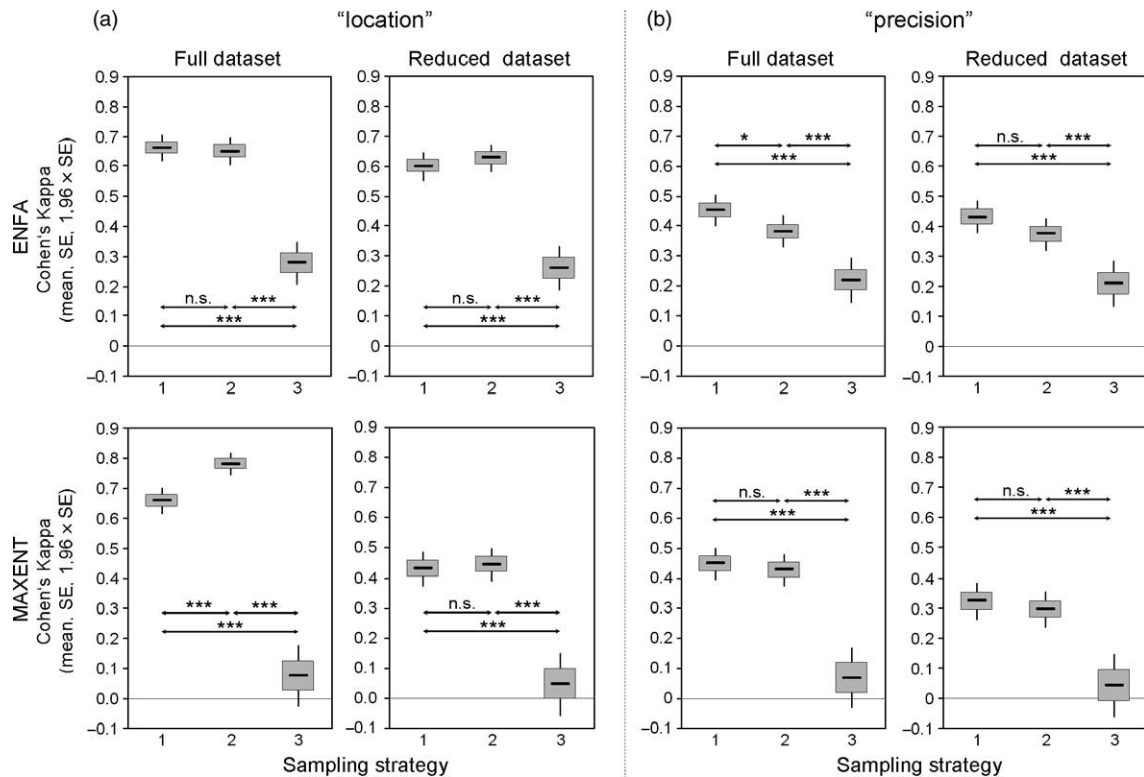


Figure 6. Accuracy of ENFA and Maxent models predicting capercaillie presence and absence, with absence data collected outside (a, “location”) and within (b, “precision”) a 1 km buffer around the presence data. Three sampling strategies are compared: (1) volunteer-based monitoring data, precise locations, (2) as (1) but data aggregated to a 1 km² grid, (3) systematically sampled data from a small subarea. Models were generated using the full dataset ($n_{(1,2)} = 587$, $n_{(3)} = 1\ 108$, and a reduced number of presence points $n_{(1,2,3)} = 60$). Cohen’s Kappa (κ) is indicated. Asterisks indicate the level of significance (** $p < 0.001$, * $p < 0.01$, * $p < 0.05$). Differences between all possible pairs of κ -values are given in Supplementary material (Table S2).

capercaillie several different types of habitat features may be required within a certain surrounding. In this case, the use of precise locations in SDM may imply pseudo-accuracy, and must not necessarily result in a substantial model improvement than when adapting the resolution to an average individual home range. Downscaling the resolution can even be advantageous to account for a required habitat mosaic (Jaberg and Guisan 2001). This can explain the good performance of the low-resolution model (sampling

strategy 2) and is supported by Graf et al. (2005) who found that a spatial resolution corresponding to a small home range best explained capercaillie distribution in the Swiss Alps.

2) Sampling extent: interpolation vs extrapolation

Animal conservation often concerns large areas and so do the requirements for model predictions. When data

Table 2. The models’ accuracy to predict capercaillie presence and absence, with absence data collected outside (A, “location”) and within (B, “precision”) a 1 km buffer around the presence data. The maximum Kappa (κ_{\max}) and the area under the receiver operating characteristics curve (AUC) are provided. Differences between κ_{\max} and mean κ_{thres} , (i.e. the Kappa value obtained when applying a threshold according to the criteria described in Supplementary material Fig. S1) are presented. Deviations between κ_{\max} and $\kappa_{\text{thres}} > 50\%$ are highlighted.

Model	Sample size	Data set	(A) “location”				(B) “precision”			
			κ_{thres}	κ_{\max}	diff (%)	AUC	κ_{thres}	κ_{\max}	diff (%)	AUC
ENFA	full	1	0.660	0.777	15.06	0.946	0.452	0.470	3.90	0.792
		2	0.650	0.752	13.56	0.943	0.382	0.408	6.45	0.765
		3	0.277	0.645	57.11	0.816	0.220	0.442	50.23	0.752
	reduced	1	0.600	0.763	21.36	0.917	0.430	0.467	7.92	0.783
		2	0.627	0.745	15.88	0.908	0.373	0.402	7.13	0.744
		3	0.260	0.573	54.62	0.765	0.208	0.400	47.92	0.710
Maxent	full	1	0.658	0.833	20.97	0.952	0.448	0.498	9.97	0.813
		2	0.780	0.842	7.36	0.968	0.428	0.432	0.85	0.767
		3	0.075	0.332	77.41	0.666	0.068	0.227	69.90	0.617
	reduced	1	0.430	0.823	47.75	0.939	0.322	0.470	31.56	0.793
		2	0.445	0.857	48.07	0.959	0.293	0.435	32.57	0.774
		3	0.047	0.485	90.38	0.744	0.042	0.337	87.64	0.673

Table 3. Effect of extrapolation on the predictive power of the strategy-3 models calibrated in the subarea. We compare (a) full extrapolations to the whole study area, (b) extrapolations restricted to the environmental domain of the subarea, (c) extrapolations with the potentially biasing variable SCVAL omitted. The capability to predict capercaillie presence and absence is evaluated using κ_{\max} and AUC (incl. SE), with absence data collected outside (A, "location") and within (B, "precision") a 1 km buffer around the presence data. Bold values indicate the mean deviance of the strategy 3 models from the two VMS-based models.

Data set		K_max (SE)						AUC (SE)					
		a		b		c		a		b		c	
A "location"													
ENFA	1	0.777	(0.018)	0.644	(0.025)	0.743	(0.019)	0.946	(0.006)	0.887	(0.011)	0.933	(0.007)
	2	0.752	(0.019)	0.604	(0.027)	0.697	(0.020)	0.943	(0.006)	0.883	(0.011)	0.921	(0.008)
	3	0.645	(0.021)	0.609	(0.026)	0.385	(0.021)	0.816	(0.012)	0.863	(0.012)	0.722	(0.014)
Maxent	1	-0.12		-0.01		-0.34		-0.13		-0.02		-0.21	
	2	0.833	(0.016)	0.734	(0.022)	0.822	(0.016)	0.952	(0.006)	0.929	(0.008)	0.964	(0.005)
	3	0.842	(0.016)	0.734	(0.022)	0.818	(0.017)	0.968	(0.004)	0.934	(0.008)	0.970	(0.004)
		0.332	(0.020)	0.304	(0.023)	0.293	(0.020)	0.666	(0.010)	0.885	(0.011)	0.905	(0.009)
		-0.51		-0.43		-0.53		-0.29		-0.05		-0.06	
B "precision"													
ENFA	1	0.470	(0.025)	0.420	(0.030)	0.460	(0.026)	0.792	(0.013)	0.766	(0.016)	0.782	(0.013)
	2	0.408	(0.025)	0.358	(0.030)	0.388	(0.026)	0.765	(0.014)	0.737	(0.016)	0.751	(0.014)
	3	0.442	(0.026)	0.414	(0.030)	0.288	(0.025)	0.752	(0.014)	0.756	(0.016)	0.697	(0.015)
Maxent	1	0.00		0.02		-0.14		-0.03		0.00		-0.07	
	2	0.498	(0.025)	0.449	(0.029)	0.475	(0.025)	0.813	(0.012)	0.788	(0.015)	0.814	(0.012)
	3	0.432	(0.026)	0.379	(0.030)	0.417	(0.026)	0.767	(0.014)	0.751	(0.016)	0.756	(0.014)
		0.227	(0.023)	0.215	(0.026)	0.225	(0.023)	0.617	(0.013)	0.736	(0.016)	0.749	(0.014)
		-0.24		-0.20		-0.22		-0.17		-0.03		-0.04	

Table 4. Performance of models in predicting the relative frequency of species presence at a local scale. Spearman rank correlations between modelled HSI scores (10 classes) and recorded observation density ($n \text{ ha}^{-1}$) in the subarea are presented, comparing models based on VMS-data with exact locations (1) and locations aggregated to a 1 km² grid (2). Two modelling methods (ENFA and MAXENT) and two different sample sizes (“full”: $n = 587$ and “reduced”: $n = 60$) are compared. No significant differences between Spearman’s R’s were recorded.

Method	Data set	Sample size	Spearman’s R	t (N-2)	Level of significance (p-value)
ENFA	1	full	0.915	6.421	0.001
	2	full	0.770	3.410	0.01
	1	reduced	0.927	7.005	0.001
	2	reduced	0.745	3.163	0.05
Maxent	1	full	0.960	9.761	0.001
	2	full	0.976	12.550	0.001
	1	reduced	0.988	18.000	0.001
	2	reduced	0.818	4.025	0.01

collection is spatially limited, extrapolations to areas outside the sampling region are required. The predictive power of extrapolated SDMs is often poor, however (Guisan and Zimmermann 2000, Graf et al. 2006). First, when the range of environmental conditions in the calibration area covers only a small part of the species’ fundamental niche, extrapolations of the species response curve beyond this range can lead to incorrect predictions (Austin et al. 1994, Thuiller et al. 2004). Second, species-habitat relationships can vary greatly between regions (Wiens et al. 1987) and spurious correlations to environmental conditions not present in the extrapolation area or predictors that are substitutable by others may lead to underestimations of the species distribution (Guisan and Zimmermann 2000). This is particularly problematic when using surrogate variables with no direct ecological significance for the species – often necessary in large-scale models, as digital maps of direct predictors (e.g. vegetation data) are mostly not available over large expanses. In fact, our models’ predictive power was mainly determined by differences in the sampling extent and the required amount of extrapolation. The precise, systematically collected data performed best within their small calibration area (Fig. 4), which agrees with the findings of Edwards et al. (2006), but predicted poorly over the whole study area. Here, capercaillie presence was most accurately predicted based on the area-wide VMS-data, which only required spatial interpolation.

The generality of a model can essentially be improved by increasing the variation in the data while capturing the broadest possible range of environmental conditions defining the species’ niche (Pearson and Dawson 2003, Thuiller et al. 2004, Guisan and Thuiller 2005) and extrapolations are most successful when the new region is similar to the calibration area (Whittingham et al. 2003, Graf et al. 2006). The latter aspect is particularly important when using presence-only approaches, as there are no absence data to contrast the presence data with. Because they discriminate between “presence” vs “random” or “used” vs “available”, these models (and their extrapolations) are not only sensitive to the range of conditions at the species locations, but also to the conditions available (Pearce and Boyce 2006).

Although chosen so as to maximise representativeness, the environmental conditions in our subarea inherently covered only a part of the variable-ranges present in the study area. For the most important predictors (i.e.

SNOWD, AGALL, FOALL, SCVAL, Table 1) these proportions ranged between 98% (SNOWD) and 56% (AGALL). In addition, for most variables there was a lower observed variance (Fig. 2). By restricting our models to the environmental range of the subarea the discrepancies between interpolated and extrapolated models could be reduced, although the latter never reached the predictive power of the strategy 1 models, maybe because the areas still differed with regard to the frequency distributions of used and available conditions along the variable gradients.

Differences of this kind were most pronounced for the two most important variables climate (SNOWD) and soil conditions (SCVAL). The latter was reflected in the model structure, as SCVAL proved much more important in the subarea than in the study area. SCVAL is an important surrogate for the presence of bilberry *Vaccinium myrtillus* (Braunisch and Suchant 2008), a key source of forage for capercaillie (Storch 1993). However, soil conditions can be partly compensated for by climate conditions, which is the case in the southern part of the study area (Braunisch and Suchant 2007), where capercaillie presence was under-predicted. Omitting this biasing but nevertheless important predictor always enlarged the areas of predicted capercaillie presence in the south (results not shown), and partly also reduced the qualitative differences between extrapolated and interpolated models, but this benefit was usually linked to a reduced overall predictive power.

Threshold selection

Converting continuous predictions into dichotomous maps – necessary for decision making in many applied situations – involves the risk of introducing a bias in model predictions (Altman et al. 1994, Fielding and Bell 1997). Clues for threshold selection should be chosen according to the conservation objectives (Wilson et al. 2005). For example, if the aim is to capture all occurrences of rare or endangered species, irrespective of the additional effort associated with unsuccessful sampling in unoccupied areas, the threshold should be chosen so as to minimise the omission error, while tolerating a high commission error – but vice versa if the goal is to optimise the investment of limited resources. However, as without true absence data specificity and commission error cannot be calculated, threshold-selection methods for presence-only data are targeted at optimising the discrimination between

predicted presence and random. The two approaches employed are both based on this principle but operate in a different way: the ROC/AUC-approach compares presence locations with a representative number of 10 000 random background cells (Phillips et al. 2006), which is possible as ROC is not sensitive to prevalence, i.e. unbalanced samples (Manel et al. 2001). Whereas sensitivity (i.e. the fraction of correctly predicted presences) can still be interpreted the classical way, the commission error (1-specificity) now corresponds to the fraction of random cells predicted as “presence”. The threshold that balances the two error types thus maximises the fraction of correctly predicted presences (sensitivity) by simultaneously minimising the total area classified as “presence”. The Continuous Boyce curve, by contrast, directly compares the area-adjusted frequency of correctly predicted presences (P) with those expected from a random distribution (E). The threshold is thus more intuitive in its interpretation as the HS-value at which P/E significantly exceeds 1, which would correspond to random prediction (cf. Hirzel et al. 2006, Supplementary material Fig. S1a). As P/E is calculated within a moving HS-window, the curve does not explicitly evaluate the effect of threshold-setting on binary predictions but rather the consistency of the continuous predictions along the HS-gradient. It therefore facilitates a reclassification into more than two classes (for details see Hirzel et al. 2006). However, small sample sizes or a low predictive power can result in rugged HS-curves with large confidence intervals which may complicate the threshold selection. It may be worth further investigating whether and to what extent the differences in the thresholding methods affect binary model predictions and contribute to differences between Maxent and ENFA. Yet, we assume such effects to be negligible for our results, as the two methods provided threshold-values in a similar order of magnitude (results not shown), and the outcome of the threshold-dependent and threshold-independent evaluation were consistent between approaches (Table 2).

The comparison of κ_{thres} and κ_{max} indicates that both approaches led to effective predictions in the VMS-data models (1 and 2), but contributed significantly to the poor model performance associated with strategy 3, reducing κ by > 50%. Clues for threshold selection based on the “use-availability” pattern observed in the calibration area may not be equally valid for the extrapolation area and can thus cause an additional decrease in model accuracy. A similar effect was found by Manel et al. (2001) who showed that locally optimised thresholds can lead to erroneous predictions when applied to areas that differ with regard to species’ prevalence.

ENFA vs Maxent

ENFA and Maxent produced models with a similar degree of accuracy when based on VMS-data (1 and 2) although there was a consistent trend in Maxent to perform slightly better, with the resulting maps – which were more compact – apparently better matching the actual capercaillie distribution. ENFA, by contrast, additionally assigned species presence to small patches in the surrounding areas. Maxent models generated from the systematically sampled data (3) performed extremely well within their

area of calibration. However, they were significantly inferior to ENFA models when extrapolated to the whole study area. Due to their great flexibility and accuracy in fitting complex response functions Maxent models are considered to be superior to envelope-based approaches (Elith et al. 2006, Phillips et al. 2006). To limit overfitting, regularisation functions are applied. These, however, refer to the empirical probability distributions of species presence along the predictor gradients in the calibration area. Consequently, regularisation may optimise the balance between complexity and parsimony within the range of calibration (Elith et al. 2006), but the close model-fit may be disadvantageous when extrapolating to areas with diverging use-availability patterns. Altering the regularisation parameters (we used the defaults) may thus be useful to explore the potentially beneficial effects of stronger regularisation. ENFA, in contrast, infers the HS-score of each location from its situation relative to the species’ optimum in the environmental space (Hirzel et al. 2002). As this is done by subdividing the environmental gradients into equal-interval classes, calculating a locations’ distance to the optimum from the sum of all species locations that lie either in the same class or in any class farther apart from the optimum, the resulting HS-curves always decrease steadily on both sides of the optimum. This interpolation between classes may level out spurious outliers in observed species-habitat relations.

We could confirm no significant advantages of Maxent when used with reduced sample sets although it performed slightly better when predicting relative observation densities. However, sixty presence points may still be too many to bring out differences. Hernandez et al. (2006) determined Maxent to be superior to Bioclim (Busby 1991), Domain (Carpenter et al. 1993) and GARP (Stockwell and Peters 1999) when reducing the sample size stepwise from 100 down to 5 records. No direct comparison with ENFA exists as yet.

Application in species monitoring and conservation

We compared three sampling strategies that required an approximately similar budget. The results show that volunteer-based survey schemes (VMS) can represent a cost-effective method of collecting species data that, due to their high degree of representativeness, can outperform data sampled locally and systematically when used for large-scale species distribution models. In the case of capercaillie, this superiority even persisted when the number of records and the spatial resolution decreased. To achieve equally good predictions with extrapolations based on local sampling, more a-priori knowledge is required, ideally based on independent evaluation data, which are mostly not available in conservation practice. Of course, we expect an increased number of subareas for systematic sampling, evenly distributed over the whole study area, to improve results for strategy 3-models considerably; however, this would necessitate substantial additional funding and manpower. Nevertheless, if the budget allows a choice between VMS-sampling and systematic surveys covering the same extent, the latter is the preferable option (see also Edwards et al. 2006). SDMs, in turn, can be a powerful tool to compensate for lacking

or spatially biased information obtained from VMS programmes, for instance by indicating areas where additional sampling may be worthwhile. By predicting relative observation densities in areas where such surveys can just prove species presence, SDMs can support the spatial prioritisation in conservation planning. Finally, periodic comparisons of SDMs from different monitoring periods enable detecting changes in species-habitat relationships, which may reveal factors responsible for population declines. However, SDMs can only support but not replace monitoring as their predictions – based on broad species-environment associations – can be obliterated by a variety of factors, such as historical factors (e.g. local extinctions), dispersal limitations, interspecific interactions (e.g. competition) or population dynamic processes (Pulliam 2000).

Conclusions

Data used for large-scale species distribution modelling should be sampled so as to best represent the range of environmental conditions in the area of interest. In presence-only models which contrast use with availability, this includes not only species locations but also available environmental conditions. In this context, models based on volunteer-based monitoring data collected haphazardly across a large, representative region can outperform models generated from systematically sampled data from a small, restricted subarea, as the latter models can suffer from a high omission error when extrapolated to areas with diverging environmental conditions and/or different “use-availability” patterns. Converting the predicted values to a binary (presence-absence) map involves the risk of introducing additional errors, as the thresholds drawn from information obtained within a restricted area may not be transferable. In the case of a highly mobile vertebrate like capercaillie, a reduction of the sample resolution to the scale of an individual home range size may not necessarily lead to a decline in model accuracy. Consequently, when restricted financial resources require choosing between data precision and spatial extent and representativeness of the sampling area, the latter strategy is less prone to prediction errors and is thus the preferable option.

Acknowledgements – We would like to thank Alexandre Hirzel and Fitsum Abadi Gebreselassie for technical advice and programming support, Karl-Eugen Schroth for fieldwork assistance, Frank Armbruster and Alexander Krämer for the development and maintenance of the monitoring database, and the numerous foresters, hunters, ornithologists and other volunteers for continuously supporting the capercaillie monitoring with data and knowledge. Raphaël Arlettaz, Phoebe Zarnetske and two anonymous reviewers provided helpful comments on the manuscript. Veronika Braunisch was financed by the Baden-Württemberg fund for environmental research.

References

Altman, D. G. et al. 1994. Dangers of using ‘optimal’ cutpoints in the evaluation of prognostic factors. – *J. Nat. Cancer Inst.* 86: 829–835.

- Austin, M. P. and Heyligers, P. C. 1991. New approach to vegetation survey design: gradsect sampling. – In: Margules, C. R. and Austin, M. P. (eds), *Nature conservation: cost effective biological surveys and data analysis*. CSIRO, pp. 31–36.
- Austin, M. P. et al. 1994. Determining species response functions to an environmental gradient by means of a beta-function. – *J. Veg. Sci.* 5: 215–228.
- Box, G. E. P. and Cox, D. R. 1964. An analysis of transformation. – *R. Stat.* 26: 211–243.
- Boyce, M. S. et al. 2002. Evaluating resource selection functions. – *Ecol. Model.* 157: 281–300.
- Braunisch, V. and Suchant, R. 2006. Das Raufußhühner – Bestandesmonitoring der FVA. – *Berichte Freiburger Forstliche Forschung* 64: 55–67.
- Braunisch, V. and Suchant, R. 2007. A model for evaluating the ‘habitat potential’ of a landscape for capercaillie *Tetrao urogallus*: a tool for conservation planning. – *Wildl. Biol.* 13: 21–33.
- Braunisch, V. and Suchant, R. 2008. Using ecological forest site mapping for long-term habitat suitability assessments in wildlife conservation – demonstrated for capercaillie (*Tetrao urogallus*). – *For. Ecol. Manage.* 256: 1209–1221.
- Braunisch, V. et al. 2008. Living on the edge – modelling habitat suitability for species at the edge of their fundamental niche. – *Ecol. Model.* 214: 153–167.
- Brotons, L. et al. 2007. Updating bird species distribution at large spatial scales: applications of habitat modelling to data from long-term monitoring programs. – *Divers. Distrib.* 13: 276–288.
- Busby, J. R. 1991. BIOCLIM – a bioclimate analysis and prediction system. – In: Margules, C. R. and Austin, M. P. (eds), *Nature conservation: cost effective biological surveys and data analysis*. CSIRO, pp. 64–68.
- Carpenter, G. et al. 1993. DOMAIN: a flexible modelling procedure for mapping potential distribution of plants, animals. – *Biodivers. Conserv.* 2: 667–680.
- Cohen, J. 1960. A coefficient of agreement for nominal scales. – *Educ. Psychol. Measur.* 20: 37–46.
- Dormann, C. F. 2007. Promising the future? Global change projections of species distributions. – *Basic Appl. Ecol.* 8: 387–397.
- Eastman, J. R. 1990. IDRISI – a grid-based geographic analysis system. – Clark Univ.
- Edwards, T. C. et al. 2006. Effects of sample survey design on the accuracy of classification tree models in species distribution models. – *Ecol. Model.* 199: 132–141.
- Elith, J. et al. 2006. Novel methods improve prediction of species’ distributions from occurrence data. – *Ecography* 29: 129–151.
- Engler, R. et al. 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. – *J. Appl. Ecol.* 41: 163–274.
- ESRI 1996. ArcView GIS 3.3. – Environmental Systems Research Inst.
- Fielding, A. H. and Bell, J. F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. – *Environ. Conserv.* 24: 38–49.
- Fleiss, J. L. 1981. *Statistical methods for rates and proportions*. – Wiley.
- Freeman, E. and Moisen, G. 2008. PresenceAbsence: an R package for presence absence analysis. – *J. Stat. Software* 23: 1–31.
- Graf, R. F. et al. 2005. The importance of spatial scale in habitat models: capercaillie in the Swiss Alps. – *Landscape Ecol.* 20: 703–717.
- Graf, R. F. et al. 2006. On the generality of habitat suitability models: a case study of capercaillie in three Swiss regions. – *Ecography* 29: 319–328.

- Guisan, A. and Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. – *Ecol. Model.* 135: 147–186.
- Guisan, A. and Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. – *Ecol. Lett.* 8: 993–1009.
- Guisan, A. et al. 1998. Predicting the potential distribution of plant species in an alpine environment. – *J. Veg. Sci.* 9: 65–74.
- Guisan, A. et al. 2005. Using niche-based models to improve the sampling of rare species. – *Conserv. Biol.* 20: 501–511.
- Guisan, A. et al. 2007. Sensitivity of predictive species distribution models to change in grain size. – *Divers. Distrib.* 13: 332–340.
- Hernandez, P. A. et al. 2006. The effect of sample size and species characteristics on performance of species distribution modeling methods. – *Ecography* 29: 773–785.
- Hirzel, A. H. et al. 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? – *Ecology* 83: 2027–2036.
- Hirzel, A. H. et al. 2006. Evaluating the ability of habitat suitability models to predict species presences. – *Ecol. Model.* 199: 142–152.
- Hirzel, A. H. et al. 2007. Biomapper 4.0. – Lab. of Conservation Biology, Dept of Ecology and Evolution, Univ. of Lausanne, Switzerland.
- Hosmer, D. W. J. and Lemeshow, S. 2000. Applied logistic regression, 2nd ed. – Wiley.
- Jaberg, C. and Guisan, A. 2001. Modelling the distribution of bats in relation to landscape structure in a temperate mountain environment. – *J. Appl. Ecol.* 38: 1169–1181.
- Jaynes, E. T. 1957. Information theory and statistical mechanics. – *Phys. Rev.* 106: 620–630.
- Jenness, J. and Wynne, J. J. 2006. Cohen's Kappa and classification table metrics 2.1: an ArcView 3x extension for accuracy of spatially-explicit models. – <www.jennessent.com/arcview/kappa_stats.htm>.
- Johnson, C. J. et al. 2006. Resource selection functions based on use-availability data: theoretical motivation and evaluation methods. – *J. Wildl. Manage.* 70: 347–357.
- Kéry, M. and Schmid, H. 2004. Monitoring programs need to take into account imperfect species detectability. – *Basic Appl. Ecol.* 5: 65–73.
- Landis, J. R. and Koch, G. G. 1977. The measurement of observer agreement for categorical data. – *Biometrics* 33: 159–174.
- Lele, S. R. 2009. A new method for estimation of resource selection probability function. – *J. Wildl. Manage.* 73: 122–127.
- Lele, S. R. and Keim, J. L. 2006. Weighted distributions and estimation of resource selection probability functions. – *Ecology* 87: 3021–3028.
- Manel, S. et al. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. – *J. Appl. Ecol.* 38: 921–931.
- Margoluis, R. and Salafsky, N. 1998. Measures of success: designing, managing and monitoring conservation and development projects. – Island Press.
- Mitchell, S. J. et al. 2001. Empirical modeling of cutblock edge windthrow risk on Vancouver Island, Canada, using stand level information. – *For. Ecol. Manage.* 154: 117–130.
- Pearce, J. L. and Boyce, M. S. 2006. Modelling distribution and abundance with presence-only data. – *J. Appl. Ecol.* 43: 405–412.
- Pearson, R. G. and Dawson, T. P. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? – *Global Ecol. Biogeogr.* 12: 361–371.
- Phillips, S. J. and Dudik, M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. – *Ecography* 31: 161–175.
- Phillips, S. J. et al. 2004. A maximum entropy approach to species distribution modeling. – In: 21st International Conference on Machine Learning. ACM Press, pp. 655–662.
- Phillips, S. J. et al. 2006. Maximum entropy modeling of species geographic distributions. – *Ecol. Model.* 190: 231–259.
- Pulliam, H. R. 2000. On the relationship between niche and distribution. – *Ecol. Lett.* 3: 349–361.
- Schneider, C. and Schönbein, J. 2003. Klimatologische Analyse der Schneesicherheit und Beschneibarkeit von Wintersportgebieten in deutschen Mittelgebirgen. – Gutachten für die Stiftung Sicherheit im Skisport des Deutschen Skiverbandes, in German.
- Schreuder, H. T. et al. 2001. For what applications can probability and nonprobability sampling be used? – *Environ. Monit. Assess.* 66: 281–291.
- Singh, N. J. et al. 2009. Using habitat suitability models to sample rare species in high-altitude ecosystems: a case study with Tibetan argali. – *Biodivers. Conserv.* 18: 2893–2908.
- StatSoft 2006. STATISTICA for Windows. – StatSoft.
- Stem, C. et al. 2005. Monitoring and evaluation in conservation: a review of trends and approaches. – *Conserv. Biol.* 19: 295–309.
- Stockwell, D. R. B. and Peters, D. P. 1999. The GARP modelling system: problems and solutions to automated spatial prediction. – *Int. J. Geogr. Inform. Syst.* 13: 143–158.
- Storch, I. 1993. Habitat selection by capercaillie in summer and autumn – is bilberry important? – *Oecologia* 95: 257–265.
- Storch, I. 1995. Annual home ranges and spacing patterns of capercaillie in central Europe. – *J. Wildl. Manage.* 59: 392–400.
- Storch, I. 2000. Conservation status and threats to grouse worldwide: an overview. – *Wildl. Biol.* 6: 195–204.
- Suchant, R. and Braunisch, V. 2004. Multidimensional habitat modelling in forest management – a case study using capercaillie in the Black Forest, Germany. – *Ecol. Bull.* 51: 455–649.
- Suchant, R. et al. 2005. Monitoring seltener Tierarten im Schwarzwald. – *Berichte Freiburger Forstliche Forschung* 58: 113–127.
- Thiel, D. et al. 2008. Ski tourism affects habitat use and evokes a physiological stress response in capercaillie *Tetrao urogallus*: a new methodological approach. – *J. Appl. Ecol.* 45: 845–853.
- Thiel, D. K. 2007. Behavioral and physiological effects in capercaillie (*Tetrao urogallus*) caused by human disturbance. – Mathematisch-naturwissenschaftliche Fakultät, Univ. Zürich.
- Thuiller, W. et al. 2004. Effects of restricting environmental range of data to project current and future species distributions. – *Ecography* 27: 165–172.
- Whittingham, M. J. et al. 2003. Do habitat association models have any generality? Predicting skylark *Alauda arvensis* abundance in different regions of southern England. – *Ecography* 26: 521–531.
- Wiens, J. A. et al. 1987. Habitat occupancy patterns of North American shrubsteppe birds: the effects of spatial scale. – *Oikos* 48: 132–147.
- Wiens, J. A. 2002. Predicting species occurrences: progress, problems, and prospects. – In: Scott, J. M. et al. (eds), Predicting species occurrences: issues of accuracy and scale. Island Press, pp. 739–749.
- Wiley, E. O. et al. 2003. Niche modelling and geographic range predictions in the marine environment using a machine-learning algorithm. – *Oceanography* 16: 120–127.

Wilson, K. A. et al. 2005. Sensitivity of conservation planning to different approaches to using predicted species distribution data. – *Biol. Conserv.* 122: 99–112.

Zarnetske, P. L. et al. 2007. Habitat classification modeling with incomplete data: pushing the habitat envelope. – *Ecol. Appl.* 17: 1714–1726.

Zweig, M. H. and Campbell, G. 1993. Receiver operating characteristic (ROC) plots – a fundamental evaluation tool in clinical medicine. – *Clin. Chem.* 39: 561–577.

Download the Supplementary material as file E5891 from www.oikos.ekol.lu.se/appendix.