

Effects of functional traits on the prediction accuracy of species richness models

Damaris Zurell¹*, Niklaus E. Zimmermann^{1,2}, Thomas Sattler³, Michael P. Nobis¹ and Boris Schröder^{4,5}

¹Dynamic Macroecology, Landscape Dynamics, Swiss Federal Research Institute WSL, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland, ²Department of Environmental Systems Science, Swiss Federal Institute of Technology ETH, CH-8092 Zürich, Switzerland, ³Swiss Ornithological Institute, Seerose 1, CH-6204 Sempach, Switzerland, ⁴Environmental Systems Analysis, Institute of Geoecology, Technische Universität Braunschweig, Langer Kamp 19c, D-38106 Braunschweig, Germany, ⁵Berlin-Brandenburg Institute of Advanced Biodiversity Research (BBIB), D-14195 Berlin, Germany

ABSTRACT

Aim We assess main ecological determinants affecting the comparative performance of macroecological models (MEMs) that model species richness directly, and stacked species distribution models based on stacking probabilities (pS-SDMs) and binary predictions (bS-SDMs). Specifically, we aimed to understand how statistical effects such as prevalence and environmental heterogeneity are entangled with species' ecology in Swiss avian assemblages.

Location Switzerland.

Methods We tested for statistical and ecological effects on overprediction and underprediction by regressing species richness residuals against communityaveraged values of prevalence, functional traits and functional dissimilarity. Further, we defined bird functional groups through hierarchical clustering and compared accuracy of species richness predictions between groups to understand the differences between model types and ecological determinants thereof. Last, we tested how accuracy of species assemblages constructed from bS-SDMs relates to species' functional characteristics.

Results Underprediction of high diversity sites by pS-SDMs and MEMs was mainly explained by prevalence, whereas overprediction of low diversity sites was strongly affected by diet and habitat traits, and increased with functional dissimilarity. Model performances varied strongly between functional groups with more accurate and less biased predictions for generalist species groups. Critically, overprediction bias in richness predictions by bS-SDMs was uncorrelated with assemblage prediction success.

Main conclusions The reliability of all community models tested here strongly depended on functional species' characteristics related mainly to diet, foraging and breeding habitat. This underlines the need to incorporate all relevant and species-specific or functional group-specific ecological filters in the models. Improved prediction accuracy of species richness will require finer-resolved environmental predictors that better describe available niche space and account for specific spatial and resource requirements of different species. More research is needed to understand the relationship between accuracy of species richness and species assemblage predictions in bS-SDMs as well as the role of biotic interactions.

Keywords

birds, community ecology, macroecological models, prediction bias, species assemblage, species richness, stacked species distribution models.

A Journal of Conservation Biogeography

*Correspondence: Damaris Zurell, Dynamic

Macroecology, Landscape Dynamics, Swiss

Zürcherstrasse 111, CH-8903 Birmensdorf,

Federal Research Institute WSL,

E-mail: damaris.zurell@wsl.ch

Switzerland.

INTRODUCTION

Species richness is one of the most important and the most widely used biodiversity measure in basic and applied ecology. Different approaches exist for modelling species richness ranging from (semi-)mechanistic (Reu et al., 2011; Mokany et al., 2012) to more empirical approaches (Ferrier & Guisan, 2006; Guisan & Rahbek, 2011). Clearly, the majority of studies rely on statistical methods because these are more easily applicable for large numbers of species and with little or no information on species' biology. Generally, we may distinguish (1) community-level and (2) species-level approaches for predicting species richness (Ferrier & Guisan, 2006). (1) Macroecological models (MEMs) directly model community properties such as species richness from a set of predictors (Ferrier & Guisan, 2006; Guisan & Rahbek, 2011). (2) In stacked species distribution models (S-SDMs), we first predict the distribution of single species using statistical species distribution models (Guisan & Zimmermann, 2000) and then combine these single-species predictions to a community-level estimate (Dubuis et al., 2011; Guisan & Rahbek, 2011). Two major stacking procedures exist, namely stacking through summing probabilities (pS-SDM) and stacking through summing binary predictions (bS-SDM). An advantage of bS-SDMs is that they directly yield local species assemblage predictions, which is not possible using MEMs

and pS-SDMs. Unfortunately, however, most previous studies point to lower prediction accuracy in bS-SDMs (Dubuis *et al.*, 2011; D'Amen *et al.*, 2015b).

MEMs and pS-SDMs tend to behave and perform very similar (Dubuis et al., 2011; Calabrese et al., 2014). Typical patterns are overestimation at sites of low species richness and underestimation at sites of high species richness. Thus, although they are able to correctly predict mean species richness and generally show symmetric prediction errors, their predictions are biased at high and low diversity sites (Fig. 1a, prediction bias I; Calabrese et al., 2014). By contrast, bS-SDMs tend towards overpredicting species richness, which is associated with generally higher and asymmetric prediction errors compared to MEMs (Fig. 1a, prediction bias II; Dubuis et al., 2011; Calabrese et al., 2014), and may be affected by the choice of threshold for making binary predictions (Cord et al., 2014; D'Amen et al., 2015a). Ecological reasons for these overprediction patterns could be missing information on biotic interactions, dispersal limitations and historical factors (Guisan & Rahbek, 2011). However, Calabrese et al. (2014) demonstrated by means of probability theory that bS-SDMs should quite generally be biased towards overprediction, independent of ecological effects.

To our knowledge, no study has yet attempted to understand and disentangle statistical and ecological effects on prediction biases in S-SDMs and MEMs (Fig. 1). Potential



Figure 1 Schematic representation of components of prediction accuracy in species richness models (a), different statistical modelling approaches for predicting species richness (b), and statistical and ecological factors affecting prediction accuracy (c). We investigate ecological and statistical effects on prediction bias I by analysing species richness residuals across communities, and *prediction bias II* by analysing prediction error (NRMSE) in different functional groups. ¹Reducible through adequate predictor choice. ²Non reducible, intrinsic property of species. ³Potentially affected by observer bias.

rather statistical effects explaining the prediction bias I (i.e. overestimating low species richness and underestimating high species richness) in pS-SDMs and MEMs are (1) prevalence, as well as (2) environmental heterogeneity and regression dilution (Calabrese et al., 2014). First, highly abundant and very rare species are difficult to model in both SDMs (McPherson et al., 2004; Santika, 2011) and MEMs (Jetz & Rahbek, 2002; Lennon et al., 2004) because of problems in identifying their dominant environmental determinants. Second, environmental heterogeneity may cause problems in identifying the specific habitat preference of species or in estimating the actual effect strength of predictors (McInerny & Purves, 2011; Calabrese et al., 2014). This may arise from inadequate resolution of the environmental predictors (regression dilution), or because limited sets of environmental predictors used do not adequately capture the ecology of all species simultaneously (Steinmann et al., 2009).

Such effects may be entangled with ecological characteristics of the modelled species, and their magnitude may differ between different species or functional groups that show particular adaptations to environmental constraints as manifested by their physiology and life history characteristics (Fig. 1c). Additionally, biotic interactions among species may greatly influence local community assemblage and should thus affect the accuracy of species richness predictions. For example, using a functional group approach, Steinmann et al. (2009) showed that MEMs of trees showed higher accuracy than MEMs of shrubs or herbs, which the authors ascribed to the high competitive ability of trees. Furthermore, it is not clear what drives the often-reported systematic overprediction in bS-SDMs (prediction bias II) compared to pS-SDMs and MEMs. Dubuis et al. (2011), Guisan & Rahbek (2011) suggested that systematic overprediction in bS-SDM arises, at least in part, from biotic interactions that restrict species co-occurrence, which was recently supported by simulated data (Thuiller et al., 2015). Additionally, the importance of biotic interactions may differ between functional groups. For example, generalist species may avoid interspecific competition using resources from different parts of their fundamental niche (Colwell & Fuentes, 1975; Martin et al., 2004). We, thus, hypothesize that bS-SDMs perform more similarly to pS-SDMs and do not exhibit systematic (or show less) overprediction for species groups with comparably weak biotic interactions, such as for example generalist species. As an effect, we expect bS-SDMs to predict species assemblages more accurately in such situations.

Here, we quantify statistical and ecological effects on prediction biases in S-SDMs and MEMs using a functional, trait-based approach (Fig. 1). Thereby, the main focus lies on species richness prediction. However, because the main advantage of bS-SDMs is the ability to construct species lists, we also assess the assemblage prediction success in bS-SDMs. Specifically:

1. We evaluate statistical and ecological effects on *prediction* bias I (Fig. 1a) by studying spatial patterns of species

richness residuals and assess the effects of threshold choice, environmental filtering, community mean prevalence and community mean traits. By this, we try to understand whether *prediction bias I* may be explained by species characteristics that, for example, relate to subscale habitat requirements.

2. We compare prediction errors of the different model types to assess the magnitude of *prediction bias II* in bS-SDMs. To this end, we model all species together or separate them into functional groups to disentangle the main ecological determinants of prediction error in species richness predictions. By this, we aim at identifying groups of species that are easier or more difficult to model by MEMs and S-SDMs and to understand the reasons behind these differences.

3. We repeat steps 1 and 2 for the analysis of species assemblage prediction success in bS-SDMs to understand the effects of species ecological characteristics on the ability of bS-SDMs to construct meaningful species lists.

This study is, to our knowledge, the first to rigorously test for and disentangle the relative importance of statistical and ecological effects on species richness models. However, such insights are prerequisite for deriving guidelines for the adequate use of S-SDMs and MEMs in predicting species richness and community assemblages. As study system, we use Swiss breeding birds, for which comprehensive and highquality data on both distribution (Schmid *et al.*, 1998) and environmental (including habitat) variables exist (Zurell *et al.*, 2012). Also, detailed information on species' biology, especially habitat and diet preferences, is available (Swiss Ornithological Institute, Pearman *et al.*, 2014).

METHODS

Species and environmental data

Presence–absence data of birds at a resolution of 1×1 km were obtained from the Swiss Breeding Bird Atlas recorded 1993–1996 (Schmid *et al.*, 1998). Within each cell, breeding birds were recorded in usually three visits (two above the tree line) with a simplified territory mapping method. Overall, approximately 90% of all species present were detected within 1 km² (Kéry & Schmid, 2006). Because very rare species are difficult to model using single-species SDMs (Breiner *et al.*, 2015), we included only species with a prevalence of at least 5% in the data (n = 104 species of the original 173 species; 2709 cells). This step was necessary to not bias the underlying SDMs unfairly and thus not bias the comparison of S-SDMs and MEMs.

Predictor variables including climatic, topographic and habitat factors were also obtained at a resolution of 1×1 km. Climatic data were compiled by the Swiss Federal Research Institute WSL (Zimmermann & Kienast, 1999) and included 11 predictors: long-term averages from the period 1961–1990 on mean summer moisture index (precipitation – potential evapotranspiration), growing degree days above 0 °C and annual values as well as values for summer and

winter for the variables precipitation sum, mean temperature and potential solar radiation (Table S1 in Supporting Information). Topographic data were derived from a digital terrain model developed by the Swiss Federal Statistical Office and included also 11 predictors: elevation (minimum, median, maximum and range), slope, aspect (sine, cosine, Beers index; Beers *et al.*, 1966) as well as secondary terrain attributes, such as the stream power index and the topographic wetness index. Land use and vegetation data were compiled from Swiss land use statistics 1992–1997 (GEOSTAT). Overall, we obtained 14 land use categories aggregated from the 74 basic categories (Table S1).

Trait data

Species' ecological traits were obtained from the Swiss Ornithological Institute and from Pearman *et al.* (2014). The five trait groups describe the trophic niche (food type, acquisition behaviour and substrate from which food is taken), habitat niche (foraging and breeding habitat), nesting position, migratory status and body size (Table S2). All traits except body size were binary. Most traits describing the trophic and the habitat niche were non-exclusive (species could have preference for several food types), except nesting position and migratory status, which were exclusive. Logtransformed body size was used as surrogate for species' space and resource requirements and was continuous.

Model fitting

We used 70% of the distribution data for model training, leaving 30% randomly selected holdout data for evaluation. Prior to modelling, all predictor variables were standardized. To reduce multicollinearity problems, in each model, we retained only predictor variables with bivariate Spearman correlations $r \leq 0.7$ giving preference to variables that were more meaningful to the species in terms of AIC (Dormann et al., 2013). Single-species SDMs were fitted with a binomial error distribution and logit link function. MEMs were fitted with a Poisson error distribution and log link. We used three different algorithms: generalized additive models GAM, multivariate adaptive regression splines MARS (Elith & Leathwick, 2007) and boosted regression trees BRT (Elith et al., 2008). GAMs were fitted with nonparametric cubic smoothing splines with up to four degrees of freedom. We estimated BRTs with a tree complexity of 2, a bag fraction of 75% and a variable learning rate such that 1000-3000 trees were fitted. The spatial autocorrelation of the residuals of SDMs and MEMs was assessed with a permutation test for Moran's I statistic with 1000 randomizations and spline correlograms, both computed in a radius of up to 20 km (Dormann et al., 2007).

Species richness predictions

We used three different methods for predicting species richness. 1) MEMs yield direct predictions of species richness. 2)

In pS-SDMs, the probabilities of occurrence predicted by SDMs were summed for all sites. 3) In bS-SDMs, we first converted the probabilistic SDM output into presence–absence predictions and subsequently summed the presences for all sites. Because the threshold choice has been reported to affect *prediction bias II* (Cord *et al.*, 2014), we used three similarly robust thresholds: species-specific sampling prevalence, TSS-maximization and sensitivity-specificity equality thresholds (Liu *et al.*, 2005).

Species richness predictions were validated against the 30% holdout data. (1) We calculated the overall deviation of predicted from observed data as the normalized root mean square error, which is

NRMSE =
$$\frac{\sqrt{\Sigma (\text{predicted - observed})^2/n}}{(\max(\text{observed}) - (\min \text{ observed}))}$$

where predicted and observed stand for the plot-by-plot observed and predicted species richness, and the range of observed biodiversity among all plots being used to standardize the errors. NRMSE served as proxy for *prediction bias II* because systematic overprediction will inflate NRMSE (Fig. 1a).

(2) In bS-SDMs, we additionally evaluated assemblage prediction success for each site, the proportion of correct predictions, which is derived from a confusion matrix (Pottier *et al.*, 2013)

Assemblage prediction success
$$=$$
 $\frac{a+d}{N}$

with a being the sum of all species correctly predicted as present and d the sum of all species correctly predicted as absent and N the total number of species.

Bias and trait analysis

First, we investigated species richness residuals and assessed to what extent community mean prevalence, community mean traits and mean functional dissimilarity can explain overestimation and underestimation of local richness in MEMs and S-SDMs. Thereby, mean functional dissimilarity served as indicator for the strength of the environmental filter. To this end, functional distances in trait space were calculated by a mixed-variable coefficient that generalizes Gower's metric of distance (Pavoine et al., 2009). A null model randomizing the functional distances among species while controlling for species richness was used to calculate the standard effect size MFD_{SES} for each community which is the rank of the observed mean functional distance within the null distribution (n = 999) divided by n + 1. MFD_{SES} varies between 0 and 1 with 0 indicating perfectly similar species and thus environmental filtering and 1 indicating completely dissimilar species and thus limiting similarity (Weiher & Keddy, 1999). Then, we regressed community mean prevalence, mean traits and MFD_{SES} against all species richness residuals. Next, we separately analysed positive and negative residuals to distinguish between factors mainly causing overprediction from those causing underprediction. We evaluated the independent effects using hierarchical partitioning (Chevan & Sutherland, 1991). Because of the high number of traits used, we applied hierarchical partitioning to models using groups of traits instead of single-trait predictors. Analogously, we regressed per-site assemblage prediction success of bS-SDMs against community mean prevalence, mean traits and MFD_{SES}.

Second, we evaluated relative model performance of MEMs and S-SDMs for separate bird functional groups. Species were grouped by a hierarchical cluster analysis using species pairwise functional distances. An average linkage strategy was used for clustering, which best approximated the original dissimilarities. Correlation between the original distance matrix and the cophenetic distance matrix from the dendrogram was assessed using the Mantel statistic (r = 0.7726; Pvalue <0.0001 with 9999 randomizations). The dendrogram was pruned to form groups such that each group contained at least five species resulting in seven functional groups (Fig. S1, Tables S3 and S4). Then, MEMs and S-SDMs were calculated for each functional group and compared with respect to NRMSE. For bS-SDMs, we additionally calculated mean assemblage prediction success (i.e. assemblage prediction success averaged over all sites).

For linking model performance to functional group characteristics, we estimated simple single-trait linear regressions and adjusted significance levels according to Holm's method (Holm, 1979). Additional to group mean traits, we calculated niche sizes for each functional group and each trait category, for example dietary niche size, as proxy for relative specialization of different groups. To this end, we ran principal component analyses for the separate trait categories (except morphology, nesting position and migratory status as these groups contained only a single trait or exclusive traits) and calculated niche size for each functional group and trait category as the area of the inertia (covariance) ellipse that describes the point cloud covered by each functional group within the first two principal axes.

All analyses were carried out within the free environment for statistical computing R (R Development Core Team, 2014).

RESULTS

SDMs and MEMs showed fair to excellent discriminatory power and good calibration (Fig. S2). Variable selection and variable importance in single-species SDMs and MEMs differed slightly between algorithms and methods with stronger differences among predictor sets of increasing complexity (Fig. S3). Spatial autocorrelation was generally low although often significant with Moran's *I* ranging between 0 and 0.15 with mean of 0.06. Spline correlograms showed mean spatial dependence of 3.3 km, but were not significant (Fig. S4). Overall, we consider the dataset as valid for the S-SDM and MEM comparison.

General patterns

As expected, MEMs and pS-SDMs yielded similar predictions with overprediction of low and underprediction of high species richness (Fig. 2a,b) whereas bS-SDM predictions exhibited general overprediction, although overprediction decreased towards high species richness (Fig. 2c). Species richness showed a decreasing relationship with elevation (Fig. 2d-f). This general pattern was well captured by all methods although bS-SDMs generally overpredicted species richness, whereas mean predicted species richness of pS-SDMs and MEMs showed almost perfect correspondence with mean observed species richness along elevation.

The systematic overprediction in bS-SDMs was reflected by NRMSE, with pS-SDMs and MEMs performing very similarly with low overall error (NRMSE = 0.11) and higher errors for bS-SDMs (NRMSE = 0.18–0.21). The different thresholds to convert probabilities into binary predictions for bS-SDMs yielded small but significant differences in NRMSE. The sensitivity–specificity equality threshold showed lowest NRMSE (and, thus, lowest *prediction bias II*) and was subsequently used for all further analyses. Algorithmic choice had almost no effect, and, in the following, we present results based on simple average consensus.

Effect of prevalence and traits on species richness residuals

Mean functional distance and community-averaged prevalence and traits explained up to 42% of residual variation in species richness predictions with very similar effects on MEMs, pS-SDMs and bS-SDMs (Table S5). To explain what drives overprediction and underprediction (*prediction bias I*), we separately analysed effects on positive and negative residuals. Because of the strong overprediction and thus low amount of negative residuals in bS-SDMs, such comparison was not meaningful for bS-SDMs because of the overriding effect of *prediction bias II*. Hence, we only report the effects on richness overprediction (explained variance 37–38%) and underprediction (explained variance 31–35%) in MEMs and pS-SDM below, which were largely congruent.

Prevalence had no effect on richness overprediction in pS-SDMs and MEMs, while it well explained underprediction, with a consistently positive effect (Fig. 3, Table S5). Thus, underprediction in communities with more wide-ranging species was reduced. Additionally, foraging and breeding habitat as well as diet explained variation in underprediction patterns. Specifically, underprediction was more severe in communities breeding in banks, coniferous forests and urban areas or foraging in dry grasslands. Patterns in overprediction were largely explained by traits related to diet, foraging and breeding habitat with minor effects from nesting position and mean functional distance. Here, overprediction was more severe in places with more carnivores, especially those that feed on other large birds, and for bird communities breeding near habitat edges or in shrubs. Furthermore, overprediction was



Figure 2 Predicted species richness against observed richness (a-c) and along elevation (d-f) for MEMs (a, d), pS-SDMs (b, e) and bS-SDMs (c, f). Shown are predictions (grey dots) along with calibration line (dashed). In panels d-e, observed species richness is depicted in black; mean predicted and observed species richness along elevation were estimated with quadratic linear models. Predictions were made on hold-out data using simple average consensus.

higher in communities exhibiting higher functional dissimilarity (resulting from competitive exclusion).

Relative model performance for bird functional groups

For bS-SDMs, we found high variability in NRMSE among functional groups and in mean assemblage prediction success (Figs 4 and 5). By contrast, NRMSE was consistently low for MEM and pS-SDM predictions with lowest errors for groups 1 (sedentary, common forest birds, partially carnivorous and omnivorous; cf. Table S3) and 7 (sedentary species breeding in tree holes). Lowest NRMSE in bS-SDMs was found for groups 1, 3 (short-distance migrants breeding on elevated nests), 5 (short-distance migrants, cavity breeders in banks and on rock faces) and 7. NRMSE in bS-SDMs was considerable higher for groups 2 (long-distance migrants), 4 (migratory ground breeders of forests and grasslands) and 6 (ground breeders of lowland marshes and highland habitat edges). For groups 1, 3 and 5, differences in NRMSE between bS-SDMs and pS-SDM/MEMs were lowest, and predictions of mean species richness by bS-SDMs were rather similar to mean species richness predicted by pS-SDMs and MEMs (Figs 4, 5 and S5). Indeed, the reduction in overprediction by bS-SDMs was significant at least for group 3 as indicated by null model testing (obtained by randomizing group memberships 5000 times while keeping number of species per group constant; Figs 4 and S11). Still, NRMSE and assemblage prediction success

were not significantly improved, and bS-SDM predictions still biased to some extent (Fig. S11).

Single-trait linear regressions had high explanatory power and explained up to 82% of the variance in NRMSE. Trait effects on NRMSE were largely congruent for MEMs and pS-SDMs while differing for bS-SDMs (Fig. 6) although the effect directions were largely the same (Table S6). For MEMs and pS-SDMs, only few traits remained significant after Holm's adjustment; for example, NRMSE was significantly reduced in bird groups with larger breeding habitat niches. Also, acquisition behaviours with active searching for food items (e.g. probing and overturning objects) significantly reduced NRMSE and may indicate a stronger success rate when foraging. For bS-SDMs, more traits remained significant. Additional to the effects reported above, which bS-SDMs share with the other model types, NRMSE in bS-SDMs was significantly lower in bird groups breeding and foraging in gardens and mixed forests and was significantly higher in groups foraging and/or breeding in or near lotic water, reed marshes, sandy or gravel sites and dry grasslands.

We found no significant correlation between single-SDM performance and S-SDM/MEM performances (significance tested against null distribution of 1000 *t*-values from randomizing group memberships of species while keeping number of species per group constant). Yet, the permutation test showed that significant relationships between prediction errors (NRMSE) and species richness or prevalence may also arise from statistical artefacts (Figs S6 and S7). For example,



Figure 3 Independent effects of different trait categories on overall explained deviance in species richness residuals of MEMs and pS-SDMs, and in assemblage prediction success in bS-SDMs. Independent contributions were calculated using a modified version of hierarchical partitioning that allows defining variable combinations. Predictions were made to hold-out data using simple average consensus. Please see Table S5 for full results of multivariate linear models.

NRMSE of pS-SDM and MEM seems to generally decrease with species richness, whereas this artefact was much less pronounced in bS-SDMs.

Assemblage prediction success

Mean assemblage prediction success over all sites for bS-SDMs was 0.8. Community mean traits explained 42% of the variance across sites with largest independent contributions from diet, foraging and breeding habitat and lower prediction success in places with large carnivores, near habitat edges and shrubland (Fig. 3, Table S5). Community mean prevalence had a smaller, consistently positive effect on assemblage prediction success. Mean functional distance had a small negative effect indicating that species assemblages were more accurately predicted in communities with more similar species (resulting from environmental filtering).

Single-trait linear regressions explained up to 63% of the variance in assemblage prediction success between functional groups (Fig. 6; Table S6). Among groups, mean assemblage prediction success in bS-SDMs significantly improved for bird groups feeding underwater or on rocky slopes and significantly decreased for bird groups foraging in canopy, and open and low forests, and for long-distance migrants breeding on elevated nests. Counter-intuitively, NRMSE and mean assemblage prediction success in bS-SDMs were not related and species assemblage was best predicted for groups 5 and 6, the latter of which exhibited highest NRMSE (Fig. 5).

DISCUSSION

This study provides the first thorough analysis of species' ecological effects on MEM and S-SDM predictions. While previous studies have hypothesized that environmental heterogeneity and prevalence effects may typically lead to overpredicting low species richness and underpredicting high species richness by MEMs and pS-SDMs (Calabrese et al., 2014), these factors have not been rigorously tested and have not been related to species' functional characteristics. Furthermore, the typical richness overprediction by bS-SDMs has been ascribed alternatively to statistical bias (Calabrese et al., 2014), to unaccounted biotic interactions (Dubuis et al., 2011; Guisan & Rahbek, 2011; Thuiller et al., 2015) or to inadequate representation of relevant environmental filters (D'Amen et al., 2015a), but has never been explicitly related to species' ecology. Using a comprehensive dataset on avian communities, we show that typical biases in richness predictions by MEMs and differently stacked S-SDMs seem to be strongly affected by species' habitat and resource requirements, which determine the degree of environmental heterogeneity needed and perceived by the species and which can partially be amended by adequate predictor choice and resolution. Prevalence only affected richness underprediction. Higher functional dissimilarity (as indicator of competitive exclusion) significantly increased richness overprediction in MEMs and pS-SDMs and significantly decreased assemblage prediction success in bS-SDMs. When modelling species richness separately for different functional groups, model performance was highly variable among groups with more accurate predictions for habitat generalists, especially for groups with large breeding habitat tolerance. Functional group analyses also showed that bS-SDMs do not generally overpredict species richness and may yield similar predictions to the other approaches for some groups. However, this is not necessarily related to more accurate predictions of



Figure 4 Predicted and observed species richness along elevation for different species groups (cf. Table S3; group 1: sedentary, common forest birds, partially carnivorous and omnivorous; group 2: long-distance migrants; group 3: short-distance migrants breeding on elevated nests; group 4: migratory ground breeders of forests and grasslands; group 5: short-distance migrants, cavity breeders in banks and on rock faces; group 6: ground breeders of lowland marshes and highland habitat edges; group 7: sedentary species breeding in tree holes). Observed species richness values are shown in black and their mean as solid black lines. pS-SDM predictions are presented by grey crosses and dashed lines (mostly not discernible from black solid lines indicating mean observed species richness), bS-SDM predictions by orange dots and solid orange lines. We tested for significantly reduced overprediction in bS-SDMs, indicated by pb, by comparing the mean difference in richness predictions and observations against a null model distribution obtained by randomly permuting group memberships 5000 times. MEMs are not shown as they generally coincide with pS-SDMs (cf. Fig. S7). Predictions were made on hold-out data using simple average consensus.

species assemblages by bS-SDMs. In the following, we discuss these issues more carefully and identify potential challenges and perspectives for predicting species richness and species assemblages.

Prediction bias intertwined with functional traits

MEMs and pS-SDM accurately predict the mean response of species richness along elevation (Fig. 2d,e; Dubuis et al., 2011; Guisan & Rahbek, 2011). Still, predictions are biased at high and low diversity sites (prediction bias I, Fig. 1a), which implies that stochastic components representing local variability are not well captured. Calabrese et al. (2014) suggested removing this bias type I by adjusting the per-species occurrence probabilities as a function of site-specific species richness. However, if we aim to project biodiversity response to environmental change, it seems more advisable to better understand the underlying causes of bias to tackle these adequately. Our results indicate that accuracy of species richness predictions is affected more strongly by environmental heterogeneity than by prevalence and is highly entangled with species' ecology, especially with habitat and resource requirements (D'Amen et al., 2015a).



Figure 5 MEM and S-SDM performance for different bird functional groups (see Fig. 4 for explanation of functional species groups). Overall prediction error is assessed as NRMSE (left). Assemblage prediction success of bS-SDMs (right) represents the mean success over all sites. Predictions were made on hold-out data using simple average consensus.

It seems promising to reduce biases in species richness predictions by further improving the environmental predictors, which corroborates recent findings for other taxa (Cord



Figure 6 Explained deviance of univariate linear models, which relate prediction error (NRMSE, as proxy for prediction bias II) and mean assemblage prediction success to trait averages within functional groups. Prior to modelling, the within-group average traits were standardised by the within-group standard deviation (except for niche sizes). We only present explained variance for traits that for at least one model type remained significant ($p\leq0.05$) after Holm's adjustment for multiple tests. The dotted line indicates the corresponding threshold for adjusted significance. Traits with prefix "e" refer to Eltonian traits related to trophic position and acquisition behaviour; prefix "f" indicates habitat traits related to foraging; prefix "b" indicates habitat traits related to breeding. Please see Tables S2 and S4 for more information on traits and within-group trait values, and Table S6 for full results of univariate linear models.

et al., 2014; D'Amen et al., 2015a). Clearly, more efforts should be given to improving habitat and resource-related variables, which are important for describing relative availability of adequate niche space and resources (Newbold et al., 2009; D'Amen et al., 2015a). For example, finerresolved habitat data could be achieved through remote sensing (Edwards et al., 1996; Cord et al., 2014). For animals and birds in particular, we see special need for improving predictors on structural habitat diversity. For example, in our analyses, underprediction and/or overprediction were more severe in communities breeding in banks or near habitat edges, and prediction error was higher for species groups that preferably breed on the ground. Ground breeders often show particular requirements regarding the structural habitat diversity for supporting both shelter and resources. Such structural facets were not well captured by our coarse-scale habitat predictors even though we tried to re-define habitat classes according to structural diversity. Another interesting aspect is that often effects of breeding habitats were more consistent among models or more important than foraging habitat. Thus, it is not sufficient to describe where the species may find enough resources to survive. For accurate biodiversity predictions, it seems also crucial to consider where the species is able to reproduce. Species' demography is hence an important aspect that needs to be considered (Normand et al., 2014; Thuiller et al., 2014), for example by integrating life stages into niche models (Taboada et al., 2013).

Adequate representation of potential resource levels is not an easy task. The analyses of species richness residuals indicated that lack of food availability might strongly affect overprediction patterns in space. For example, overprediction tended to be more severe in communities with higher carnivore richness. Also, assemblage prediction success in bS-SDMs was reduced for carnivorous diets, although these effects did not remain significant after Holm's adjustment. We attribute these effects mostly to the spatial requirements of birds of prey. For these species, large home ranges are typical, and these are not well captured at the 1 km resolution used here. We, thus, strongly recommend that home range sizes should be considered when deciding on spatial resolution of biodiversity models and that biodiversity models should move towards employing a variety of species or functional group-specific spatial resolutions, instead of one universal spatial resolution as is usually done. By contrast, the presence of omnivores reduced overprediction. This may be due to the opportunistic and highly adaptable behaviour of omnivores such as Corvidae that are able to utilize a wide range of habitats. Diet and behavioural traits are thus an important determinant for the accuracy of biodiversity predictions.

Prevalence is a factor with an essentially non-reducible effect on *prediction bias I* (Fig. 1). However, our analyses showed that the relative effect of prevalence is comparably low. It is only important for explaining richness

underprediction in MEMs and pS-SDMs. We found lower underprediction, that is lower prediction bias, for more prevalent species. However, community average prevalence only ranged from low to intermediate values (0.16–0.66), and, thus, our results corroborate previous findings that prediction accuracy is reduced for very narrow ranging species but potentially also for very wide-ranging species (Jetz & Rahbek, 2002; Calabrese *et al.*, 2014). Here, it is interesting to note that prevalence may also be affected by observer bias, which is generally low in the Swiss breeding bird data (Kéry & Schmid, 2006), but may prove more important in other datasets. Observer bias may inflate the importance of the prevalence effect on prediction accuracy of species richness predictions and should therefore receive more attention.

Comparative model performance and future avenues

pS-SDMs and MEMs produced very similar predictions and were also similarly affected by species ecological traits. Predictions by bS-SDMs, on the other hand, differed more strongly. General overprediction of species richness (prediction bias II; Fig. 1a) led to typically higher prediction errors (NRMSE) in bS-SDMs. Calabrese et al. (2014) argued in mathematical terms that S-SDMs based on stacking binary predictions should quite generally lead to biased results. Results from our functional group analysis indicate large differences in the magnitude of this bias. For two functional groups (group 3 and 5), the mean richness was predicted well and overprediction was significantly reduced for group 3. Still, all bS-SDM predictions had higher NRMSE, and thus were more biased, than either MEMs or pS-SDMs. These results partially corroborate findings by D'Amen et al. (2015a) that bS-SDMs may, in some cases, yield similarly accurate predictions of species richness as pS-SDMs and MEMs. Overall, NRMSE in bS-SDMs tended to be lower for habitat generalists foraging and breeding in gardens and mixed forests and higher in habitat specialists foraging and breeding in reeds, or gravel banks. This partially supports our initial hypothesis that prediction bias II in bS-SDMs may be lower for generalist species which can avoid interspecific competition using resources from different regions of their fundamental niche (Colwell & Fuentes, 1975; Martin et al., 2004). Still, it provides no conclusive evidence that overprediction in bS-SDMs is indeed driven by unaccounted biotic interactions (Dubuis et al., 2011; Guisan & Rahbek, 2011), but merely emphasizes that accuracy of species richness predictions may decrease in the face of high niche partitioning. This, again, may relate to subscale environmental heterogeneity not well captured by the predictors and underlines the strong entanglement of statistical and ecological effects (Fig. 1c). It remains open how richness overprediction in bS-SDMs relates to contemporary biotic interactions, which should certainly be explored in more detail in the future. For example, recently published joint species distribution models could be used for disentangling environmental and biotic

effects although they may also similarly suffer from unaccounted environmental heterogeneity (Pollock *et al.*, 2014; Warton *et al.*, 2015).

The main advantage of bS-SDMs is their ability to construct species lists directly and then to predict biodiversity inventories and to derive species turnover. Our results indicate that prediction success for local assemblages strongly depends on species' ecology. Critically, species richness prediction errors in bS-SDMs and assemblage prediction success were not correlated. Thus, less-biased species richness predictions do not necessarily correspond to more accurate species lists predicted from bS-SDMs. This could indicate difficulties in predicting the outcome of stochastic biotic processes such as competition, but also dispersal and disturbances among others, which requires further exploration in the future. Without further model development, we have to caution that bS-SDMs might not be capable of making good simultaneous predictions of species richness and species assemblages.

Overall, we see great potential for functional approaches to improving biodiversity predictions. First, functional groups can be identified for which simple bS-SDM predictions are sufficient for predicting assemblages and those that require more complicated community-level approaches (Mokany et al., 2011; Pollock et al., 2014; D'Amen et al., 2015a). Second, a functional approach may allow for a more targeted, trait-based selection of environmental predictor sets and adequate selection of spatial resolution. Clearly, more elaborate analyses are needed for clarifying the role of demographic processes and biotic interactions as well as sampling errors for predictions of community composition. For animals, the study of functional groups and biotic interactions will also require to consider trophic interactions. Finally, more work is needed to improve predictions for rare and specialist species as these seem to be the most difficult groups to predict while also being the most vulnerable at the same time (Platts et al., 2014).

ACKNOWLEDGEMENTS

We thank the dedicated and highly qualified volunteers who conduct the fieldwork of the Swiss breeding bird atlas, Niklaus Zbinden for early discussions on the subject and Laure Gallien for valuable comments on the manuscript. DZ received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement no 624958. NEZ acknowledges support from the Swiss National Science Foundation (SNSF) (grant: 31003A_149508/1). BS acknowledges funding by the German Science Foundation (Grants Nos. SCHR1000/6-2 and SCHR1000/8-2).

REFERENCES

Beers, T.W., Dress, P.E. & Wensel, L.C. (1966) Aspect transformation in site productivity research. *Journal of Forestry*, 64, 691–692.

- Breiner, F.T., Guisan, A., Bergamini, A. & Nobis, M.P. (2015) Overcoming limitations of modelling rare species by using ensembles of small models. *Methods in Ecology and Evolution*, 6, 1210–1218.
- Calabrese, J.M., Certain, G., Kraan, C. & Dormann, C.F. (2014) Stacking species distribution models and adjusting bias by linking them to macroecological models. *Global Ecology and Biogeography*, **23**, 99–112.
- Chevan, A. & Sutherland, M. (1991) Hierarchical partitioning. *The American Statistician*, **45**, 90–96.
- Colwell, R.K. & Fuentes, E.R. (1975) Experimental studies of the niche. *Annual Review of Ecology and Systematics*, **6**, 281–310.
- Cord, A.F., Klein, D., Gernandt, D.S., de la Rosa, J.A.P. & Dech, S. (2014) Remote sensing data can improve predictions of species richness by stacked species distribution models: a case study for Mexican pines. *Journal of Biogeography*, **41**, 736–748.
- D'Amen, M., Pradervand, J.-N. & Guisan, A. (2015a) Predicting richness and composition in mountain insect communities at high resolution: a new test of the SESAM framework. *Global Ecology and Biogeography*, **24**, 1443–1453.
- D'Amen, M., Dubuis, A., Fernandes, R.F., Pottier, J., Pellissier, L. & Guisan, A. (2015b) Using species richness and functional traits predictions to constrain assemblage predictions from stacked species distribution models. *Journal of Biogeography*, **42**, 1255–1266.
- Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Kühn, I., Ohlemüller, R., Peres-Neto, P.R., Reineking, B., Schröder, B., Schurr, F.M. & Wilson, R. (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, **30**, 609–628.
- 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**, 27–46.
- Dubuis, A., Pottier, J., Rion, V., Pellissier, L., Theurillat, J.-P. & Guisan, A. (2011) Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking modelling approaches. *Diversity and Distributions*, **17**, 1122–1131.
- Edwards, T.C. Jr, Deshler, E.T., Foster, D. & Moisen, G.G. (1996) Adequacy of wildlife habitat relation models for estimating spatial distributions of terrestrial vertebrates. *Conservation Biology*, **10**, 263–270.
- Elith, J. & Leathwick, J. (2007) Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. *Diversity and Distributions*, **13**, 265–275.
- Elith, J., Leathwick, J.R. & Hastie, T. (2008) A working guide to boosted regression trees. *Journal of Animal Ecology*, 77, 802–813.

- Ferrier, S. & Guisan, A. (2006) Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology*, **43**, 393–404.
- Guisan, A. & Rahbek, C. (2011) SESAM a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography*, **38**, 1433–1444.
- Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, 135, 147–186.
- Holm, S. (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, **6**, 65–70.
- Jetz, W. & Rahbek, C. (2002) Geographic range size and determinants of avian species richness. *Science*, **297**, 1548–1551.
- Kéry, M. & Schmid, H. (2006) Estimating species richness: calibrating a large avian monitoring programme. *Journal of Applied Ecology*, **43**, 101–110.
- Lennon, J.J., Koleff, P., Greenwood, J.J.D. & Gaston, K.J. (2004) Contribution of rarity and commonness to patterns of species richness. *Ecology Letters*, **7**, 81–87.
- Liu, C., Berry, P.M., Dawson, T.P. & Pearson, R.G. (2005) Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, **28**, 385–393.
- Martin, K., Aitken, K.E.H. & Wiebe, K.L. (2004) Nest sites and nest webs for cavity-nesting communities in interior british columbia, canada: nest characteristics and niche partitioning. *The Condor*, **106**, 5–19.
- McInerny, G.J. & Purves, D.W. (2011) Fine-scale environmental variation in species distribution modelling: regression dilution, latent variables and neighbourly advice. *Methods in Ecology and Evolution*, **2**, 248–257.
- McPherson, J.M., Jetz, W. & Rogers, D.J. (2004) The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *Journal of Applied Ecology*, **41**, 811–823.
- Mokany, K., Harwood, T.D., Overton, J.M., Barker, G.M. & Ferrier, S. (2011) Combining alpha- and beta-diversity models to fill gaps in our knowledge of biodiversity. *Ecology Letters*, **14**, 1043–1051.
- Mokany, K., Harwood, T.D., Williams, K.J. & Ferrier, S. (2012) Dynamic macroecology and the future for biodiversity. *Global Change Biology*, **18**, 3149–3159.
- Newbold, T., Gilbert, F., Zalat, S., El-Gabbas, A. & Reader, T. (2009) Climate-based models of spatial patterns of species richness in Egypt's butterfly and mammal fauna. *Journal of Biogeography*, **36**, 2085–2095.
- Normand, S., Zimmermann, N.E., Schurr, F.M. & Lischke, H. (2014) Demography as the basis for understanding and predicting range dynamics. *Ecography*, **37**, 1149–1154.
- Pavoine, S., Vallet, J., Dufour, A.-B., Gachet, S. & Daniel, H. (2009) On the challenge of treating various types of variables: application for improving the measurement of functional diversity. *Oikos*, **118**, 391–402.
- Pearman, P.B., Lavergne, S., Roquet, C., Wüest, R., Zimmermann, N.E. & Thuiller, W. (2014) Phylogenetic patterns of

D. Zurell et al.

climatic, habitat and trophic niches in a European avian assemblage. *Global Ecology and Biogeography*, 23, 414–424.

- Platts, P.J., Garcia, R.A., Hof, C., Foden, W., Hansen, L.A., Rahbek, C. & Burgess, N.D. (2014) Conservation implications of omitting narrow-ranging taxa from species distribution models, now and in the future. *Diversity and Distributions*, 20, 1307–1320.
- Pollock, L.J., Tingley, R., Norris, W.K., Golding, N., O'Hara, R.B., Parris, K.M., Vesk, P.A. & McCarthy, M.A. (2014) Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution*, 5, 397–406.
- Pottier, J., Dubuis, A., Pellissier, L., Maiorano, L., Rossier, L., Randin, C.F., Vittoz, P. & Guisan, A. (2013) The accuracy of plant assemblage prediction from species distribution models varies along environmental gradients. *Global Ecol*ogy and Biogeography, **22**, 52–63.
- R Development Core Team (2014) *R: A Language and Environment for Statistical Computing*.R Foundation for Statistical Computing, Vienna, Austria. URL https://www.Rproject.org/ (accessed 25 July 2014).
- Reu, B., Proulx, R., Bohn, K., Dyke, J.G., Kleidon, A., Pavlick, R. & Schmidtlein, S. (2011) The role of climate and plant functional trade-offs in shaping global biome and biodiversity patterns. *Global Ecology and Biogeography*, 20, 570–581.
- Santika, T. (2011) Assessing the effect of prevalence on the predictive performance of species distribution models using simulated data. *Global Ecology and Biogeography*, **20**, 181–192.
- Schmid, H., Luder, R., Naef-Daenzer, B., Graf, R. & Zbinden, N. (1998) Schweizer Brutvogelatlas. Verbreitung der Brutvögel in der Schweiz und im Fürstentum Liechtenstein 1993-1996, Swiss Ornithological Institute, Sempach, Switzerland.
- Steinmann, K., Linder, H.P. & Zimmermann, N.E. (2009) Modelling plant species richness using functional groups. *Ecological Modelling*, **220**, 962–967.
- Taboada, A., von Wehrden, H. & Assmann, T. (2013) Integrating life stages into ecological niche models: a case study on tiger beetles. *PLoS ONE*, **8**, e70038.
- Thuiller, W., Münkemüller, T., Schiffers, K.H., Georges, D., Dullinger, S., Eckhart, V.M., Edwards, T.C. Jr, Gravel, D., Kunstler, G., Merow, C., Moore, K., Piedallu, C., Vissault, S., Zimmermann, N.E., Zurell, D. & Schurr, F.M.T. (2014) Does probability of occurrence relate to population dynamics? *Ecography*, **37**, 1155–1166.
- Thuiller, W., Pollock, L.J., Gueguen, M. & Münkemüller, T. (2015) From species distributions to meta-communities. *Ecology Letters*, **18**, 1321–1328.
- Warton, D.I., Blanchet, F.G., O'Hara, R.B., Ovaskainen, O., Taskinen, S., Walker, S.C. & Hui, F.K.C. (2015) So many

variables: joint modeling in community ecology. *Trends in Ecology & Evolution*, **30**, 766–779.

- Weiher, E. & Keddy, P. (1999) *Ecological assembly rules: per-spectives, advances.* Cambride University Press, Retreats.
- Zimmermann, N.E. & Kienast, F. (1999) Predictive mapping of alpine grasslands in Switzerland: species versus community approach. *Journal of Vegetation Science*, **10**, 469–482.
- Zurell, D., Grimm, V., Rossmanith, E., Zbinden, N., Zimmermann, N.E. & Schröder, B. (2012) Uncertainty in predictions of range dynamics: Black Grouse climbing the Swiss Alps. *Ecography*, **35**, 590–603.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

 Table S1. Predictor variables used for prediction of Swiss

 avian species distributions and richness.

Table S2. Trait variables.

Table S3. Species groups as defined by hierarchical clustering.

Table S4. Average trait values within functional groups.

Table S5. Effects of ecological traits on species richness residuals and on assemblage prediction success.

Table S6. Effects of mean traits within functional groups on prediction error (NRMSE) and mean assemblage prediction success.

Figure S1. Dendrogram of functional dissimilarity and resulting bird functional groups.

Figure S2. Model performance of single SDMs and MEMs.

Figures S3-S6. Variable importance in SDMs and MEMs.

Figure S7. Spatial autocorrelation of single species SDMs and S-SDMs within bird functional groups.

Figure S8. Predicted and observed species richness along elevation for different bird functional groups.

Figure S9. Effects of mean SDM accuracy, maximum local species richness and prevalence within functional groups on bS-SDM and pS-SDM performance.

Figure S10. Null distributions for effects of SDM accuracy prevalence and richness on bS-SDM and pS-SDM performance.

Figure S11. Null distributions of bS-SDM performance (mean overprediction, NRMSE, and assemblage prediction success) for bird functional groups.

BIOSKETCH

Damaris Zurell is a post-doctoral researcher in macroecology and biodiversity modelling at the Swiss Federal Research Institute WSL. Her research combines ecological theory, statistical and mechanistic modelling approaches, and empirical data. Main interests include environmental change effects on spatio-temporal structuring of species' ranges, populations and communities, and the evolution of life history characteristics, especially in migratory species. Author contributions: D.Z., B.S. and N.E.Z. conceived the ideas; D.Z. designed the study with substantial input from all authors. T.S. advised and helped with species and trait data; N.E.Z., B.S. and M.N. advised and helped on methodology. D.Z. performed all analyses and led the writing.

Editor: Janet Franklin