

RESEARCH PAPER



Model complexity affects species distribution projections under climate change

Philipp Brun¹ | Wilfried Thuiller² | Yohann Chauvier¹ | Loïc Pellissier^{1,3} |
Rafael O. Wüest¹ | Zhiheng Wang⁴ | Niklaus E. Zimmermann¹

¹Swiss Federal Research Institute (WSL),
Birmensdorf, Switzerland

²Univ. Grenoble Alpes, Univ. Savoie Mont
Blanc, CNRS, LECA, Laboratoire d'Écologie
Alpine, Grenoble, France

³Department of Environmental Systems
Science, Landscape Ecology, Institute of
Terrestrial Ecosystems, ETH Zürich, Zürich,
Switzerland

⁴Institute of Ecology and Key Laboratory
for Earth Surface Processes of the Ministry
of Education, College of Urban and
Environmental Sciences, Peking University,
Beijing, China

Correspondence

Philipp Brun, Swiss Federal Research
Institute (WSL), 8903 Birmensdorf,
Switzerland.
Email: philipp.brun@wsl.ch

Funding information

Schweizerischer Nationalfonds zur
Förderung der Wissenschaftlichen
Forschung, Grant/Award Number:
310030L_170059; Swiss Federal Office for
the Environment; Agence Nationale de la
Recherche, Grant/Award Number: ANR-10-
LAB-56 and ANR-15-IDEX-02

Handling Editor: Damaris Zurell

Abstract

Aim: Statistical species distribution models (SDMs) are the most common tool to predict the impact of climate change on biodiversity. They can be tuned to fit relationships at various levels of complexity (defined here as parameterization complexity, number of predictors, and multicollinearity) that may co-determine whether projections to novel climatic conditions are useful or misleading. Here, we assessed how model complexity affects the performance of model extrapolations and influences projections of species ranges under future climate change.

Location: Europe.

Taxon: 34 European tree species.

Methods: We sampled three replicates of predictor sets for all combinations of 10 levels ($n = 3$ –12) of environmental variables (climate, terrain, soil) and 10 levels of multicollinearity. We used these sets for each species to fit four SDM algorithms at three levels of parameterization complexity. The >100,000 resulting SDM fits were then evaluated under environmental block cross-validation and projected to environmental conditions for 2061–2080 considering four climate models and two emission scenarios. Finally, we investigated the relationships of model design with model performance and projected distributional changes.

Results: Model complexity affected both model performance and projections of species distributional change. Fits of intermediate parameterization complexity performed best, and more complex parameterizations were associated with higher projected loss of current ranges. Model performance peaked at 10–11 variables but increasing number of variables had no consistent effect on distributional change projections. Multicollinearity had a low impact on model performance but distinctly increased projected loss of current ranges.

Main conclusions: SDM-based climate change impact assessments should be based on ensembles of projections, varying SDM algorithms as well as parameterization complexity, besides emission scenarios and climate models. The number of predictor variables should be kept reasonably small and the classical threshold of maximum absolute Pearson correlation of 0.7 restricts collinearity-driven effects in projections of species ranges.



KEYWORDS

climate change, European trees, future projection, model performance, multicollinearity, number of variables, parameterization complexity, species distribution modelling, species range change, species range loss

1 | INTRODUCTION

Efficient mitigation of biodiversity loss from global changes requires a thorough understanding of how species' ranges are organized in space, and how they will shift in the future. Two approaches are commonly employed to establish such understanding: statistical species distribution models (SDMs, Guisan & Zimmermann, 2000) and mechanistic models (e.g. Zurell et al., 2016). Projections of species range shifts using mechanistic models are based on explicitly formulated processes that are presumably relevant to the ecology of the target species, while SDM projections extrapolate relationships identified from statistical structures between occurrences and their environment. In principle, projections from mechanistic models may seem preferable as their careful application may harbour a lower risk that relevant processes are insufficiently captured or corrupted by erroneously identified associations (Merow et al., 2014). However, limited understanding of relevant processes and of the ecology of most species, and/or lack of relevant data to describe it sufficiently well prevent their use in many cases (Guisan & Zimmermann, 2000; Thuiller et al., 2008). Despite their limitations, statistical SDMs are therefore likely to remain commonly used to project species responses to global change. For this reason, it is imperative to comprehend the implications of the various conceptual decisions taken at the different steps of the development of SDM projections.

Implications of decisions in projection design can be quantified by comparing the outcomes of alternative setups when projected under climate change (aka projection ensembles). Projection ensembles consist of multiple projections generated by systematically varying the settings at the different steps of their development, such as initial conditions, that is, the presence and (pseudo)absence data used for model training, predictor variables, SDM algorithms, parameterization complexity, climate models, or emission scenarios. Projection ensembles are particularly useful to quantify uncertainty and to obtain consensus projections, which are arguably superior to single model projections (Araújo & New, 2007, but see Dormann et al., 2018). Furthermore, if combined with rigorous model validation, projection ensembles can help identifying model designs of relatively high quality. Compared to other fields, such as economics and climate science, projection ensembles were introduced to species distribution modelling relatively recently (Araújo & New, 2007; Thuiller, 2004), but gained popularity since specialized modelling platforms became available—such as the R-package 'biomod2' (Thuiller, Lafourcade, Engler, & Araújo, 2009). However, not all steps in the development of projection ensembles have received the same level of attention. A literature study of 125 recent papers employing SDM projections revealed that the most frequently varied step was the emission scenario (63% of cases), followed by the climate models

used to estimate future climatic conditions (48% of cases) (Figure 1a, for further information see Appendix S1). SDM algorithms and initial conditions were also frequently varied (35% and 32% of cases, respectively). Implications of decisions revolving around model complexity, on the other hand, were typically not explored, and either left to the defaults of the method applied or taken based on more or less well-grounded heuristics. Yet, the importance of also varying model complexity in projection ensembles has recently been emphasized by several authors (Boria, Olson, Goodman, & Anderson, 2014; Merow et al., 2014; Werkowska, Márquez, Real, & Acevedo, 2017).

Most SDM algorithms can be tuned to fit models across a substantial range of complexity, from 'under fit' models that are not flexible enough to capture the detailed species response to the environment to 'over fit' models that ascribe signal to noise, which is particularly risky when projecting (Merow et al., 2014; Moreno-Amat et al., 2015). Even when differences in model performance are minor, projections from complex models can strongly differ from those of simple models (Beaumont et al., 2016; Gregor, Palacios, Thompson, & Chan, 2018; Merow et al., 2014). However, systematically varying model complexity across different SDM algorithms is not straightforward, as their different setups do not allow for analogous tuning, and universal measures to directly compare complexity are lacking (García-Callejas & Araújo, 2016). Various proxies for model complexity have therefore been suggested, including the shapes of response curves, predictor and parameter numbers, and the computation time required for model fitting (Bell & Schlaepfer, 2016; García-Callejas & Araújo, 2016; Merow et al., 2014). We investigate the roles of three aspects related to model complexity: parameterization complexity, number of variables used, and multicollinearity among variables. Parameterization complexity involves modifications of a set of parameters, adjusting the level of complexity within SDM algorithms. These variations can be based on the flexibility of response curves or the inclusion of interaction terms in regression techniques and tree complexity in tree-based methods (Merow et al., 2014). Varying parameterization complexity has not been employed routinely in the recent literature. Among the 125 papers that we investigated, it was varied only twice (Figure 1a). Instead, algorithms were mostly run with default parameterizations or else with simplifications of the default flexibility (see also Hao, Elith, Guillera-Aroita, & Lahoz-Monfort, 2019).

Model complexity is also affected by the number of predictor variables considered as well as their multicollinearity. Adding more predictors to a model increases the amount of signal and noise available to SDM algorithms and typically leads to larger numbers of parameters estimated, and thus more complex models (Merow et al., 2014; Werkowska et al., 2017). However, many algorithms include strategies to eliminate parameters that

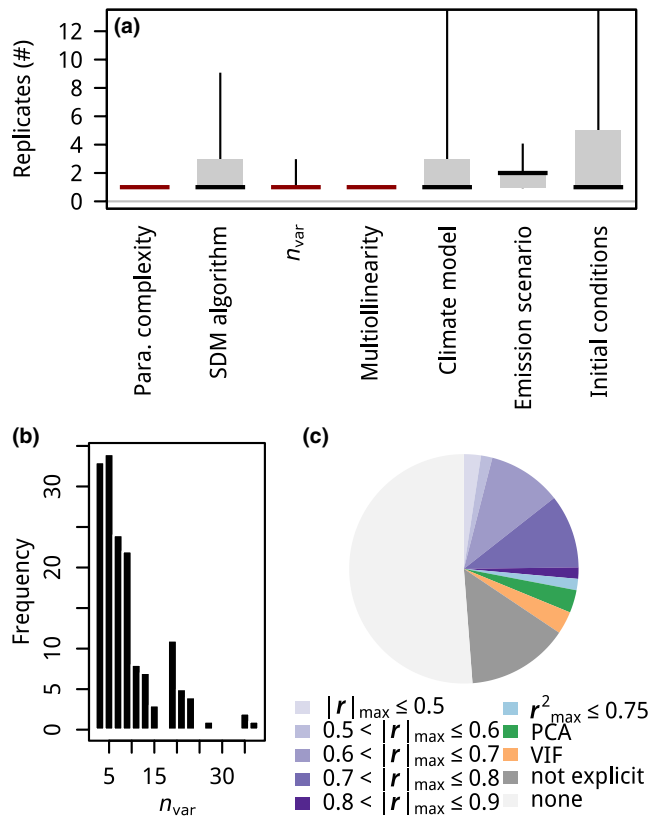


FIGURE 1 Replications and decisions for different steps in projection development in 125 recent studies using SDMs to project future distributional changes. Number of replicates for key steps in the development of SDM projections are shown in (a). Central lines represent medians, boxes represent interquartile ranges, and whiskers indicate 95% confidence intervals. For the labels n_{var} represents number of variables; 'Para. Complexity' refers to parameterization complexity; and 'initial conditions' include resampling of presence observations and pseudo-absences. (b) A histogram of the number of variables used in SDMs for projecting is shown. (c) The most common criteria used to limit multicollinearity in predictor sets. $|r|$ represents the Pearson correlation coefficient, PCA stands for principal component analysis, and VIF abbreviates variance inflation factor; 'not explicit' summarizes studies which indicate that they reduced multicollinearity without explicitly stating the criteria that were employed. For details, see Appendix S1

insufficiently improve model fits, which leads to a saturating relationship between number of variables and model complexity. Particularly many parameters may be eliminated for predictor sets with high levels of multicollinearity, and thus a limited amount of independent information. Multicollinearity may therefore lead to somewhat simpler models. But investigating the effects of multicollinearity is also of interest because it can compromise parameter estimates which are especially problematic when models are transferred to situations with different multicollinearity regimes (Dormann et al., 2013). As ecologically important predictors often show significant levels of collinearity, knowing the maximum level of tolerable collinearity is critical. Among the 125 papers we investigated, the median number of variables included was seven, ranging from two to 37 (Figure 1b).

Yet, within the same analysis the numbers were typically not varied (only in 6% of cases), and if they were, then mainly as a consequence of recombining variable groups (e.g. climate vs. climate and soil variables) and not to study the impact of numbers of variables. Also, multicollinearity levels were only exceptionally varied (2% of cases), and the heuristics used to limit multicollinearity varied greatly (Figure 1c).

In this study, we analyzed a comprehensive ensemble of SDM projections and compared uncertainty associated with the commonly varied decision steps in ensembles (SDM algorithm, emission scenario, and climate models) with uncertainty originating from parameterization complexity, number of variables, and multicollinearity. Furthermore, we investigated the patterns of model performance, projections of distributional change, and disagreement of projections of distributional change (i.e. variation from replicated predictor sets) along model complexity gradients. Using survey data for 34 tree species across Europe, we fitted and evaluated more than 100,000 SDMs with two performance metrics, and generated over 800,000 projections of species distribution ranges that we summarized with two metrics of distributional change. Based on the results of these primary analyses, we addressed the following questions:

1. Which are the most important factors affecting the performance of model extrapolations to 'novel' (non-analogous) conditions, and projections of species distributional change?
2. Are the effects of model complexity on model performance and species distributional change in line with the expectations formulated in Table 1?

2 | MATERIALS AND METHODS

2.1 | Overview

Our analyses consisted of three steps. First, we prepared a comprehensive set of environmental variable combinations. We established a pool of 24 climate variables for both, current and future conditions, and a pool of 16 soil/terrain variables which we assumed to remain constant until 2080. Based on pairwise Pearson correlation coefficients, we defined 100 combinations of numbers of variables and multicollinearity levels, and screened the realm of possible predictor sets with roughly equal numbers of climatic and soil/terrain variables for three replicates per combination. Second, we evaluated and projected a large number of SDM fits (Figure 2). For each combination of predictor set and species, we fitted four SDM algorithms at three levels of parameterization complexity and evaluated their performance. Then, we projected the fitted models to the conditions in 2061–2080 as projected by four climate models for two emission scenarios and assessed projected species distributional changes. Third, we investigated how model complexity affects model performance, projected distributional change, and disagreement between projections of distributional change.



TABLE 1 Expected and found relationships of model performance and projected species distributional change with parameterization (para.) complexity, number of variables (n_{var}), and multicollinearity

Factor	Expectation	Reason	Finding
Model performance			
Para. complexity	⤿	Models with too simple parameterizations are not flexible enough to capture the detailed species response to the environment whereas models with too complex parameterizations ascribe signal to noise, deteriorating the skill of model extrapolations. We expect optimal model performance at intermediate parameterization complexity	⤿
n_{var}	⤿	Too few predictors yield insufficiently informed occurrence-environment relationships. Adding predictors remedies this deficit but also adds noise to the data and thus increases the risk of ascribing signal to noise. We expect model performance to peak at moderate numbers of variables and to decline if predictor numbers become too high	⤿
Multicol-linearity	↘	Multicollinearity inflates the variance of regression parameters and potentially leads to the misidentification of relevant predictors (Dormann et al., 2013). We expect the performance of model extrapolations to decrease with increasing multicollinearity	—
Distributional change			
Para. complexity	↗	We expect models with more complex parameterizations to fit a tighter niche shape around conditions of occurrence than models with simpler parameterizations, due to their increased flexibility (Warren et al., 2014). From these tighter fits we expect greater changes in projected distributions	RL ↗ RC —
n_{var}	↗	Adding predictor variables provides additional criteria to constrain fits of ecological niches, and thus, by tendency, leads to more constrained niche fits. From these more constrained fits we expect greater changes in projected distributions	RL — RC —
Multicol-linearity	↗	Model fits are optimized conditional to the linear associations between predictors (multicollinearity). Such associations are likely to be different in training and projection data. Model fits trained on data with high levels of multicollinearity therefore respond to both changing future conditions and changing future associations and thus we expect them to show greater changes in projected distributions	RL ↗ RC —
Distributional change disagreement			
Para. complexity	↗	The more complex parameterizations are, the more likely they ascribe signal to noise. Consequently, we expect higher disagreement between projections to different (replicated) predictor sets of same size when modelled with more complex parameterizations	RL ⤿ RC ↘
n_{var}	↗	The more predictor variables are supplied to an algorithm, the more possibilities exist to constrain a niche. Consequently, we expect higher disagreement between projections to different (replicated) predictor sets of same size when modelled with more predictors	RL ↗ RC ↗
Multicollinearity	↗	Changing linear associations between predictors represents an additional source of uncertainty in future projections. Consequently, we expect higher disagreement between projections to different (replicated) predictor sets of same size when modelled with increasing levels of multicollinearity	RL ↗ RC ↗

Note: Distributional change represents expectations for both species range loss (RL) and species range change (RC, see Section 2). Distributional change disagreement refers variation from replicated predictor sets. Grey indicates weak relationships.

2.2 | Data

2.2.1 | Species distribution data

Our distribution data originated from the international cooperative program on the assessment and monitoring of air pollution effects on forests (ICP Forests Level 1). The ICP Forests surveys forest conditions in Europe on a grid of roughly 16 km horizontal resolution (Lorenz, 1995). From this program, we used presence/absence data collected in the years 2005–2007 containing observations at 6,146 locations in total. We constrained the study area to the European continent and the British Isles, extending to the eastern boundary of the European Union, and removed 563 observations,

mostly originating from Belarus and remote islands south of 36°N (Cyprus, Crete, Canaries, see Figure S2.1 in Appendix S2 for a map). Furthermore, we focused on 34 tree species with 50 or more presence observations and distinguished the seven species with more than 300 observations (or presence in a representative area of at least 76,800 km²) as the subgroup 'common' species. Species names and observation are listed in Table S3.2 in Appendix S3.

2.2.2 | Environmental data

All 24 climate and 16 soil/terrain variables described below were projected to the standard projection for Europe (EPSG 3035), and aggregated to 4 km horizontal resolution.

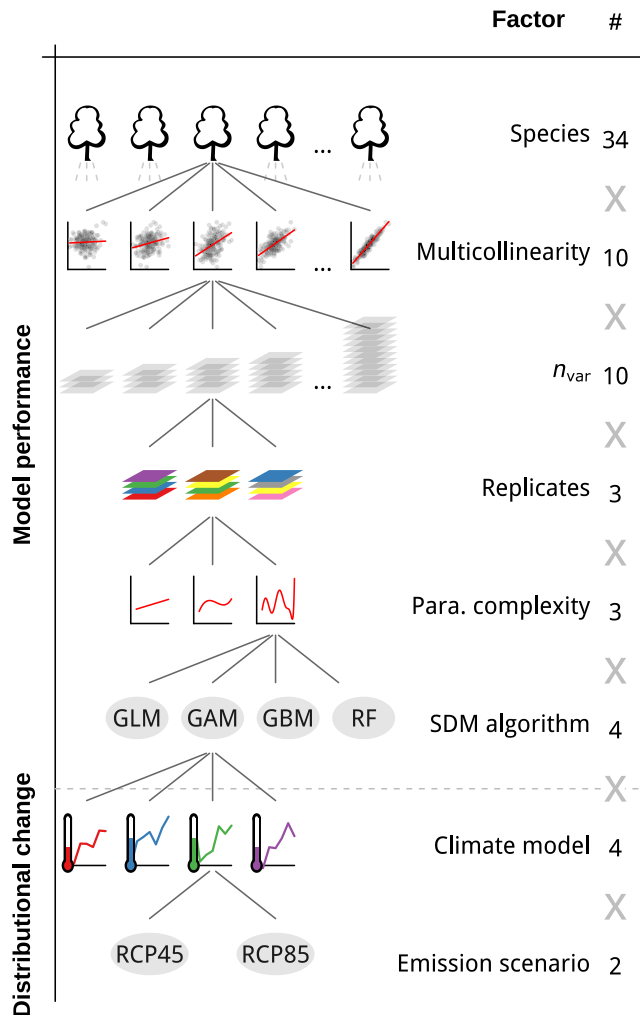


FIGURE 2 Set-up of primary analyses. Above the grey, dashed line are factors (steps in projection development), whose effects on model performance were evaluated; below the line are additional factors included for investigating drivers of range change. " n_{var} " are numbers of variables

Climate variables

The climate variables for present and future conditions were taken from the CHELSA initiative with an original resolution of 30 arc-sec (Karger et al., 2017, <http://chelsa-climate.org/>, accessed September 2018). We included annual mean temperature, isothermality, temperature seasonality, maximum temperature of the warmest month, minimum temperature of the coldest month, the annual range of temperature, annual precipitation, and precipitation seasonality. Furthermore, for each quarter we estimated mean, maximum, and minimum temperature and precipitation. For future conditions, we considered projections to 2061–2080 from the four available IPCC climate models that scored highest in skill and complementarity (Sanderson, Knutti, & Caldwell, 2015): CESM1-BGC, CMCC-CM, MIROC5, and MPI-ESM-MR, as well as two emission scenarios, that is, the representative concentration pathways 45 and 85 (RCP45 and RCP85, IPCC, 2013).

Soil/terrain variables

We considered five terrain variables and eleven soil variables. From the European Environment Agency digital elevation model (EU-DEM, accessed November 2017) with 25 m horizontal resolution, we derived terrain ruggedness (maximum elevation difference in each 4×4 km cell), mean slope, maximum difference of slope in each cell, and aspect. In addition, we used the topographic index, that is, the tendency of the soil to become saturated with water as a result of the topography of the surrounding area, with an original resolution of 15 arc-sec (Marthews, Dadson, Lehner, Abele, & Gedney, 2015). Soil variables were taken from the European Soil Data Centre. We used water content at saturation, water content at field capacity, water content at wilting point, and soil hydraulic conductivity from the maps of indicators of soil hydraulic properties for Europe with 1 km original horizontal resolution (Tóth et al., 2015, accessed November 2017). Estimates of topsoil organic carbon were taken from (Jones, Hiederer, Rusco, & Montanarella, 2005, accessed November 2017) with 1 km original horizontal resolution. Finally, from the European Soil Database we used the fractions of gravel, sand, silt, and clay, as well as rooting depth and bulk density with 1 km original horizontal resolution (Hiederer, 2013, accessed September 2018). Where information was available for topsoil and subsoil, we calculated profile averages before spatially aggregating by average. In order to have roughly normally distributed predictors, we log-transformed terrain ruggedness, maximum difference of slope, and soil organic carbon estimates.

2.3 | Analyses

2.3.1 | Generation of predictor sets

We used two criteria to define 10 levels of multicollinearity (Table 2) in order to cover a range of multicollinearity levels, and to include current best practice recommendations. We distinguished 10 bins of increasing third quartiles of absolute Pearson correlation coefficients ($|r|$). Varying third quartiles of $|r|$ allows

TABLE 2 Definition of correlation bins used in this study

Third quartile of $ r $	Fraction of $ r > 0.7$	Common interpretation
[0–0.1)	0	OK
[0.1–0.2)	0	OK
[0.2–0.3)	0	OK
[0.3–0.4)	0	OK
[0.4–0.5)	0	OK
[0.5–0.6)	0–0.33	Critical
[0.6–0.7)	0–0.33	Critical
[0.7–0.8)	>0.33	Bad
[0.8–0.9)	>0.33	Bad
[0.9–1]	>0.33	Bad

Note: $|r|$: vector of absolute pairwise Pearson correlation coefficients; 'common interpretation' refers to the frequent employment of $|r| \leq 0.7$ as an upper boundary for multicollinearity (see Figure 1).



for modifying the multicollinearity structure of the whole predictor set, rather than just constraining its extremes. Furthermore, we linked these bins to a fixed fraction of pairwise combinations that have an $|r|$ above 0.7 (Dormann et al., 2013), which facilitates the interpretation of multicollinearity levels. Then, we screened our variable pool for three replicates of all possible combinations between three to 12 variables (i.e. 10 sets of differing variable numbers) and the 10 multicollinearity levels (Table 2) under the constraint that climate and soil/terrain variables had to be represented in similar fractions (Appendix S4). The resulting 300 potential predictor set combinations allow a sound assessment of the major patterns of model performance and projected distributional change in number of variables \times multicollinearity space.

2.3.2 | Species distribution modelling

Algorithms

For each combination of species and predictor set we fitted four SDM algorithms with a simple, an intermediate, and a complex parameterization, each. The algorithms included two regression techniques, generalized linear models (GLMs, McCullagh & Nelder, 1983) and generalized additive models (GAMs, Hastie & Tibshirani, 1990), and two tree-based techniques, random forest (RF, Breiman, 2001) and gradient boosting machines (GBM, Friedman, 2001). General settings independent of parameterization complexity included the following: for the regression techniques we assumed a binomial error distribution, used the logit link function, and up-weighted presence observations to obtain a balanced prevalence of 0.5. RF fits were based on 500 trees, and in GBM we fitted trees with a complexity of five and a learning rate of 0.005 and also up-weighted presences to obtain a prevalence of 0.5. SDMs were fitted in the R environment (version 3.5.1) using the packages 'gam' (version 1.16), 'randomForest' (version 4.6-14), and 'gbm' (version 2.1.5) (Greenwell, Boehmke, Cunningham, & Developers, 2018; Hastie, 2018; Liaw & Wiener, 2002; R Development Core Team, 2008).

Parameterization complexity

Parameterization complexity mainly involved adjusting the flexibility of the response curves in regression techniques, the terminal node size in RF, and the number of trees in GBMs. Simple GLMs were fitted considering intercept and linear terms; for intermediate GLMs we added second order polynomials; and for complex GLMs we also included third and fourth order polynomials and first order interactions. For GAMs we set the degree of freedom in the smooth terms to 1.5, three, and 10 for simple, intermediate, and complex parameterizations, respectively. Simple, intermediate, and complex GBMs were distinguished by training 100, 300, and 10,000 trees, respectively. In a preliminary analysis 10,000 trees were identified to have a predictive deviance near the minimum for the chosen learning rate and the data at hand. Finally, complexity of RFs was varied by setting the minimum number of observations in the terminal nodes to 40, 20 and one in simple, intermediate, and complex RFs, respectively.

Model performance

We assessed model performance based on two metrics, True Skill Statistic (TSS, Allouche, Tsoar, & Kadmon, 2006), and area under the curve (AUC, Swets, 1988). These metrics were derived from model projections to 'novel' conditions using environmental block cross-validation (Roberts et al., 2017, Appendix S2.1 in Appendix S2). Block cross-validation is a comparably tough test enforcing projections to conditions that are somewhat more different than our future environmental conditions were from present conditions, both in terms of the covered ranges and correlation structure (Appendix S2.2 in Appendix S2).

Estimating distributional change

We estimated two metrics of distributional change, range loss and range change. Species range loss is defined as the percentage of currently suitable habitats that are expected to be lost under future conditions. Species range change is defined as the percentage of change in the entire range, assuming dispersal to all newly suitable habitats. Both measures were assessed from binary presence/absence projections that were converted from the continuous model outputs using the threshold maximizing TSS. For both measures we estimated the disagreement, that is, the variation imposed by replicated predictor sets, with interquartile ranges.

2.3.3 | Meta-analyses on model performance and distributional change

Our meta-analyses on model performance and distributional change consisted of two parts: (a) We investigated the patterns of model performance, distributional change, and distributional change disagreement along the gradients of number of variables, multicollinearity and parameterization complexity. (b) We used analysis of variance (ANOVAs) to assess the relative contributions of the steps in projection development to uncertainty in model performance and projections of distributional change. These analyses were based on the outputs of the primary analyses plus missing value imputations for number of variables \times multicollinearity combinations for which no predictor sets were available.

Analysis of patterns and missing value imputation

We summarized model performance and distributional change in the number of variables \times multicollinearity space to investigate their patterns and to generate estimates for missing values. For estimating missing values, we represented the combinations of number of variables and multicollinearity bins in a 10×10 pixel space separately for each combination of species, SDM algorithm, and parameterization complexity (plus emission scenario and climate model in the case of distributional change). Then, we pixel-wise summarized model performance and projected distributional change estimates from the three replicates by median and approximated pixels with missing values with bilinear interpolations from neighbouring pixels. To investigate patterns, we combined original data and imputed

missing values, and similarly summarized pixels by median and interquartile range (IQR) but for pooled estimates from all species.

Analysis of variance

We used ANOVA to quantify the relative contributions of the different sources of uncertainty in projection ensembles. We ran ANOVAs with model performance metrics (TSS and AUC) and with distributional change estimates (range loss, range change) as response. For model performance ANOVAs, we compared the contributions of number of variables, multicollinearity, parameterization complexity, and SDM algorithm. For distributional change ANOVAs, we additionally considered the effects of the two different climate models and the two different emission scenarios. In order to have comparable level numbers for the different factors, we aggregated number of variables and multicollinearity to two levels: low and high levels of multicollinearity were distinguished by a third quartile of $|r|$ of 0.5 while the group of low numbers of variables included 3–7 and the group of high numbers 8–12. This aggregation resulted in 75 potential predictor sets for each of the four combinations of aggregated n_{var} and multicollinearity levels. To account for non-independence resulting from the nestedness of SDM algorithm and parameterization complexity, we additionally considered their linear interaction. We accounted for species identity through a random intercept. ANOVAs were based on Bayesian generalized linear mixed models, fitted with the Integrated Nested Laplace Approximations (INLA) approach (Rue, Martino, & Chopin, 2009). Instead of p -values, which are not helpful for large sample sizes, we used parameter uncertainty in the posterior distributions to assess how distinct mean sums of squares of the different factors were. We estimated mean sums of squares 1,000 times based on resampled parameter estimates from the posterior distributions of the fitted INLA models, and report medians and 95% confidence intervals. For response variables bounded by zero and one (AUC, range loss) we assumed errors to follow a beta distribution, otherwise normal error distribution was assumed.

3 | RESULTS

3.1 | Model performance

3.1.1 | Analysis of variance

The number of variables was the most important factor explaining variations in TSS (Figure 3). With 0.71, the fraction of total sum of squares (frac_{SST}) was six times higher for number of variables than for any other factor. Second and third most important factors were parameterization complexity and SDM algorithm with similar frac_{SST} of 0.13 and 0.12, respectively. Multicollinearity, on the other hand, was the least important factor ($\text{frac}_{\text{SST}} < 0.01$). For TSS of fits for 'common' species (>300 presence observations), the ranking was the same, but parameterization complexity ($\text{frac}_{\text{SST}} = 0.19$) was distinctly more important than SDM algorithm ($\text{frac}_{\text{SST}} = 0.07$, Figure S5.3 in Appendix S5). The results of the AUC ANOVA were very similar to those of the TSS ANOVA (Figure S5.4 in Appendix S5).

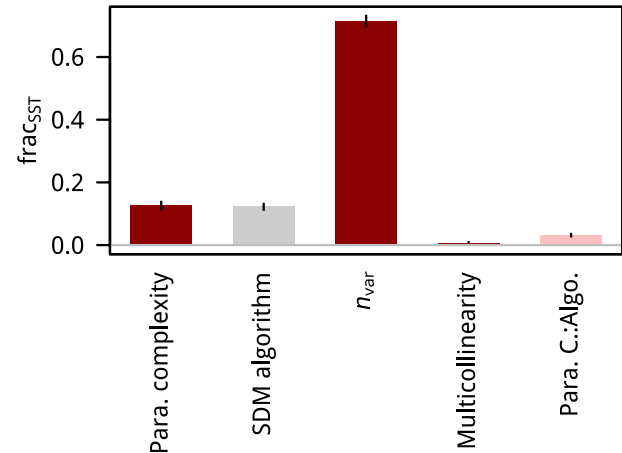


FIGURE 3 Effects of SDM modelling decisions on TSS as assessed by ANOVA. Bars represent fractions of total sums of squares between levels (frac_{SST}); error bars indicate 95% confidence intervals. Steps affecting model complexity are highlighted in dark red. The light red bar labeled 'Para. C.:Algo.' represents the interaction between parameterization complexity and SDM algorithm

3.1.2 | Analysis of patterns

Overall, TSS measured under environmentally extrapolating block cross-validation was highest for parameterizations of intermediate complexity, showed a unimodal relationship with number of variables, and no clear relationship with multicollinearity (Figure 4). TSS increase was steep for models built with three to five variables, started levelling-off for models built on >5 variables, and typically peaked at 10 or 11 variables (Figure 4b). Fits of intermediate and high complexity achieved notably higher TSS than those of low complexity. Their TSS was similar if no more than five variables were included; otherwise fits of intermediate parameterization complexity outperformed complex fits (Figure 4b). Interquartile range of TSS tended to decrease with parameterization complexity, in particular for GLMs and GAMs (Figure S5.5 in Appendix S5), indicating that under these conditions the type of predictor variable used had a comparably low impact on performance. TSS of fits of 'common' species was on average slightly lower than overall TSS, but the patterns were generally similar (Figure S5.6 in Appendix S5). However, 'common' species fits showed a weak negative relationship with multicollinearity, and among those fits complex parameterizations achieved highest TSS.

Model performance patterns remained similar when assessed by AUC, and when relative instead of absolute predictor numbers were considered. As TSS, AUC showed a unimodal relationship with number of variables with highest scores at 10 or 11 variables, it peaked for parameterizations of intermediate complexity, and showed no clear relationship with multicollinearity (Figure S5.7 in Appendix S5). Relationships with model performance were also mostly unimodal when the number of variables per presence observation rather than the absolute number of variables was considered (Figure S5.8 in Appendix S5): TSS and AUC were typically highest if one predictor was used for about 10 presence observations.

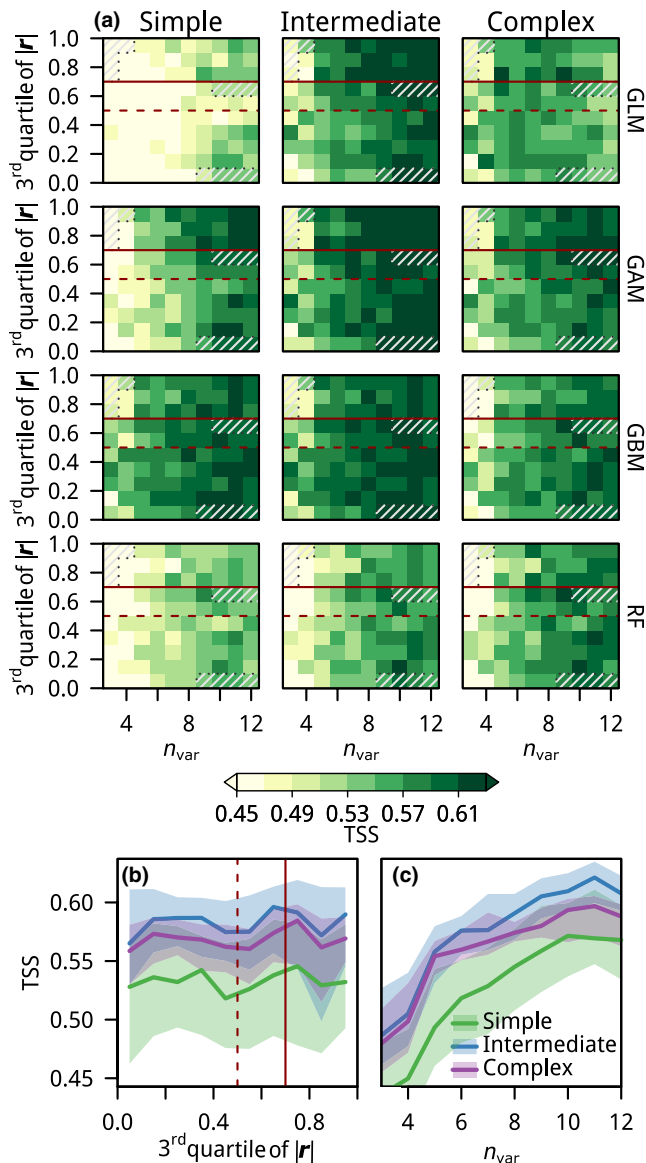


FIGURE 4 TSS patterns along gradients of number of variables, multicollinearity and parameterization complexity. (a) TSS of model fits in the number of variables \times multicollinearity space. Rows separate SDM algorithms; columns distinguish levels of parameterization complexity. Dark red lines illustrate correlation thresholds as defined in Table 2; hashed areas represent level combinations for which no predictor sets were available. (b, c) TSS scores are summarized along multicollinearity (b) and number of variables (c) gradients. Relationships for models with low, intermediate, and high parameterization complexity are shown in green, blue, and purple, respectively. Lines represent medians; polygons represent interquartile ranges

3.2 | Species distributional change

3.2.1 | Analysis of variance

Emission scenario was the most important factor explaining variations in projected range loss, but parameterization complexity was almost as important (Figure 5). In the range loss ANOVA, the

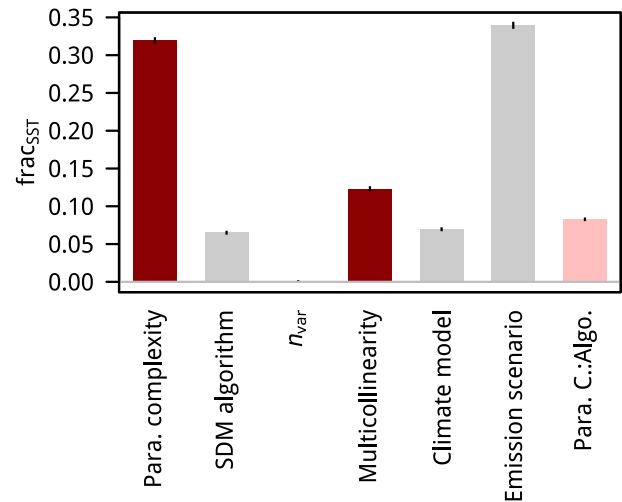


FIGURE 5 Effects of SDM modelling decisions on range loss as assessed by ANOVA. Bars represent fractions of total sums of squares between levels (frac_{SST}); error bars indicate 95% confidence intervals. Steps affecting model complexity are highlighted in dark red. The light red bar labeled 'Para. C.:Algo.' represents the interaction between parameterization complexity and SDM algorithm

proportion of total sum of squares of emission scenario and parameterization complexity was about one-third each (0.34 and 0.32, respectively). Third most important factor was multicollinearity ($\text{frac}_{\text{SST}} = 0.12$). The contributions of SDM algorithm, climate model and the interaction between parameterization complexity and SDM algorithm similarly were relatively less important (frac_{SST} between 0.06 and 0.08), and number of variables made no notable contribution. Range loss projections of 'common'-species models, on the other hand, were mainly affected by emission scenario ($\text{frac}_{\text{SST}} = 0.49$), and parameterization complexity was comparably less important: with a frac_{SST} of 0.10, it contributed less than multicollinearity ($\text{frac}_{\text{SST}} = 0.11$, Figure S6.9 Appendix S6). In contrast, variations in projections of range change were largely driven by SDM algorithm ($\text{frac}_{\text{SST}} = 0.75$, Figure S6.10 in Appendix S6).

3.2.2 | Analysis of patterns

Higher fractions of ranges were projected to be lost by fits with more complex parameterizations and predictor sets with elevated levels of multicollinearity (Figure 6). On average parameterizations of intermediate complexity projected a median range loss that was 16% higher than that of parameterizations of low complexity; fits with parameterizations of high complexity projected another 5% increase (Figure 6b,c). These differences were driven by projections of GLMs and GAMs which were particularly affected by parameterization complexity (Figure S6.11 in Appendix S6). Median projected range loss also increased by 10% for predictor sets with a third quartile of $|r|$ larger than 0.5 (Figure 6b). For 'common' species, range loss projections were on

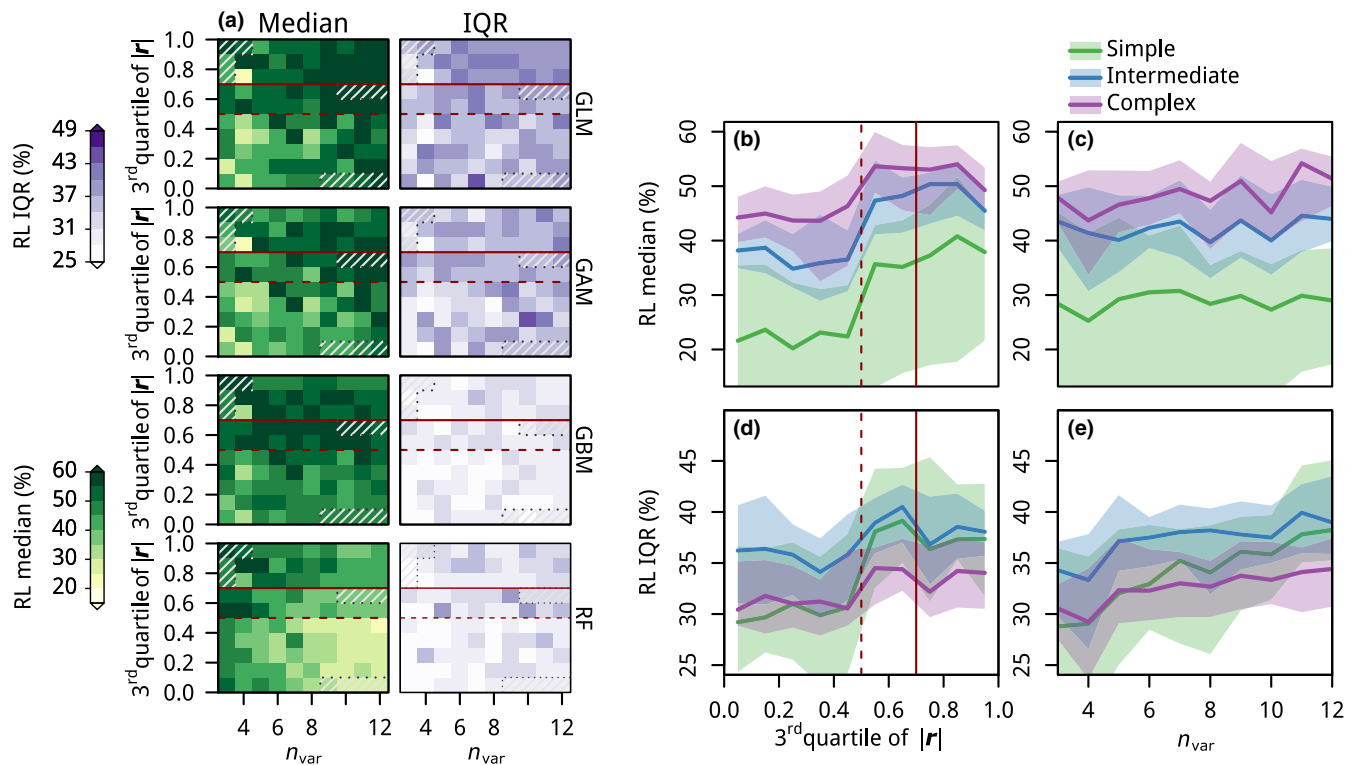


FIGURE 6 Range loss patterns along gradients of number of variables and multicollinearity. (a) Percent range loss of SDMs with parameterizations of intermediate complexity in the number of variables \times multicollinearity space. Rows separate SDM algorithms; the left column represents the median range loss, the right column represents the range loss interquartile range. Dark red lines illustrate correlation thresholds as defined in Table 2; hashed areas represent level combinations for which no predictor sets were available. (b–e) Median range loss and its interquartile range along gradients of multicollinearity (b,d) and number of variables (c, e) for models fitted with parameterizations of low (green), intermediate (blue), and high (purple) complexity. Lines represent medians; polygons represent interquartile ranges. Range loss patterns in the number of variables \times multicollinearity space for parameterizations of low and high complexity are shown in Figure S6.12 in Appendix S6

average slightly higher, with similar but weaker patterns (Figures S6.12 and S6.13 in Appendix S6). As for range loss, patterns of range change responded to parameterization complexity, with models fitted with parameterizations of low complexity projecting net range gains, and those fitted with other complexity levels projecting net range losses (Figures S6.14 and S6.15 in Appendix S6). However, SDM algorithms caused even larger differences in projected range change: RF projections estimated notable net range gains (33% on average), whereas all other algorithms overall projected little change or net range losses (3%, –5%, and –8% on average for GLM, GAM and GBM, respectively). Responses of projected distributional change were much more pronounced when relative rather than absolute numbers of variables were considered (Figure S6.16 in Appendix S6). Range loss showed a concave relationship with number of variables per presence observation with minima at between 10 and 25 presences per variable; range change projections, in contrast, peaked at these numbers.

The interquartile range of range loss projections varied considerably for different combinations of SDM algorithm and parameterization complexity, and showed a weakly positive relationship with both number of variables and multicollinearity (Figure 6).

The IQR of range loss projections was higher for GLMs and GAMs than for GBMs and RF (Figure 6 and Figure S6.11 in Appendix S6). Relationships with multicollinearity and number of variables were both increasing but rather weak, while range loss IQR was slightly higher for parameterizations of intermediate complexity than for those of high or low complexity. Patterns were similar among ‘common’-species models, although among them simple fits were associated with highest range loss IQR (Figures S6.12 and S6.13 in Appendix S6). IQR of projected range change also tended to increase with number of variables and multicollinearity, and decreased with parameterization complexity (Figures S6.14 and S6.15 in Appendix S6). Range change IQR was furthermore especially high for RF projections.

4 | DISCUSSION

The findings of our analyses suggest that varying model complexity is crucial in the development of SDM projections, to optimize model performance and to capture uncertainty in range loss projections. All assessed model performance metrics responded by far most strongly to the number of variables considered, typically showing a unimodal



relationship with optimal performance at 10–11 variables (although performance declines at high numbers of variables were modest). Furthermore, parameterization complexity contributed nearly one third to uncertainty in range loss projections, almost as much as that from the contrasting scenarios of future emissions. Multicollinearity was also important in this context: predictor sets with a third quartile of $|r|$ above 0.5 were associated with 10% higher range loss projections.

Parameterization complexity has varying implications on model performance, and increases range loss but not range change projections. As expected, we found parameterizations of intermediate complexity to yield highest TSS and AUC. These findings are in agreement with reported loss in extrapolation performance (i.e. transferability) of models fitted from comparably complex parameterizations (e.g. Chala et al., 2016; Gegr et al., 2018) and of comparably complex SDM algorithms (e.g. Randin et al., 2006). Interestingly, this relationship was not consistent across all assessments: TSS of complex parameterizations was highest when considering 'common' species (>300 presence observations). Fits with complex shapes may be more appropriate for prevalent species when rich and well-designed data are available, as was the case in this study. Furthermore, in line with our expectations fits with more complex parameterizations projected higher levels of species range loss. Tighter niche fits indeed appear to increase expected loss of suitable habitat (Warren, Cardillo, Rosauer, & Bolnick, 2014), although 'common' species seem relatively robust to this artefact (Figure S6.9 in Appendix S6). For real-world applications we would have excluded model fits with simple parameterizations from the projection ensemble as they performed clearly worst. But the performance of model fits of intermediate and high complexity was similar and thus both of these complexity levels contribute relevant information and should be considered.

A minimum number of variables is necessary to allow for well-performing models, but at a certain point adding variables no longer improves and even starts deteriorating model performance, and it increases the disagreement among projections of distributional change. Performance of model extrapolations typically was highest for models trained on 10 or 11 predictors. This relationship corresponds to the expected unimodal shape, and it is similar to findings of SDM transfers in space, which, however, often showed performance maxima for less than five predictors (Duque-Lazo, Gils, Groen, & Navarro-Cerrillo, 2016; Petitpierre, Broennimann, Kueffer, Daehler, & Guisan, 2017). Likely, these lower maxima are the consequence of greater changes of environmental conditions in spatial transfers as compared to projections under climate change, in particular when some predictors, such as soil variables, are assumed to remain constant over the time scales considered (Appendix S2.2 in Appendix S2). Furthermore, the optimal number of variables is dataset-specific. For sampling designs similar to the one we worked with, one predictor per 10 presence observations may be ideal (Harrell et al., 1998). However, for less well-designed survey data such as presence-only data, finer grains or steeper environmental gradients, more presence observations per predictor may be necessary. For very

large datasets the ratios may be even lower, as information contained in predictor variables tends to get increasingly redundant when more predictors are added, in particular if only climate variables are considered.

Multicollinearity has surprisingly little effect on model performance, but can lead to a distinct increase in range loss projections. This lack of clear negative effects on the performance of model extrapolations is different from our expectation (Table 1) and somewhat surprising, but it corresponds with results from previous studies (e.g. Petitpierre et al., 2017; Wauchope et al., 2017). Multicollinearity inflates the variance of regression parameters and potentially leads to the misidentification of relevant predictors (Dormann et al., 2013). But its impact on model performance appears to be modest even for well-sampled, prevalent species, which was the most susceptible group in this study (Figure S5.6 in Appendix S5). Perhaps, the multicollinearity gradient within typical environmental predictors is too narrow to cause substantial errors in model projections. Nevertheless, the impact of multicollinearity is large enough to markedly affect range loss estimates. Across all species, range loss projections increased steeply for predictor sets with a third quartile of $|r|$ larger than 0.5, that is, predictor sets including variable pairs with absolute Pearson correlation coefficients above the commonly used threshold of 0.7 (Dormann et al., 2013). The different levels of multicollinearity tested were not hypotheses of equivalent relevance but represent increasingly severe violations of a central model assumption: independence among predictors. Although differences in model performance are low, the increased range loss for predictor sets with maximum $|r| > 0.7$ therefore indicates that multicollinearity starts having notable consequences on projections above this threshold which should be avoided.

The architecture of SDM algorithms may be more important than parameterization complexity when it comes to model extrapolations to novel conditions. In our analyses, emission scenarios contributed most to uncertainty in projections of species range loss while SDM algorithms dominated uncertainty in range change projections. SDM algorithms have also been found to be the major drivers of uncertainty in range change projections of mammals, birds, amphibians and freshwater fish (Buisson, Thuiller, Casajus, Lek, & Grenouillet, 2010; Garcia, Burgess, Cabeza, Rahbek, & Araújo, 2012; Thuiller, Guéguen, Renaud, Karger, & Zimmermann, 2019), while emission scenarios have been shown to contribute most to uncertainty in range loss projections (Thuiller et al., 2019). Thuiller et al. (2019) argue that these differences may arise from the different extrapolation behaviors of SDM algorithms, which is particularly relevant in newly suitable habitats which are considered in range change but not in range loss projections. RF models appear to particularly promote this uncertainty by projecting range change estimates that strongly deviate from those by other algorithms (Figures S6.14–S6.16 in Appendix S6, Beaumont et al., 2016). Our results indicate that differences in extrapolation behaviour may be mainly determined by the SDM algorithm and less by the selected parameterization complexity. With the increasing availability of long-term observational and environmental data it may become possible to

constrain the group of SDM algorithms with relevant extrapolation behaviour based on empirical evidence, leaving decisions on model complexity also among the key factors affecting range change projections.

Based on our results and the considerations discussed above, we formulate three recommendations for including model complexity in ensemble simulations of climate change impact on biodiversity using SDMs:

1. SDM algorithms and parameterization complexity: SDM algorithms as well as parameterization complexity have important consequences on projected distributional change and thus both factors should be varied in ensembles. Appropriate levels of parameterization complexity depend on the dataset at hand, and can be constrained based on model performance. We suggest to run SDM algorithms at least at two levels of parameterization complexity. Under computational constraints, this may go at the cost of using many SDM algorithms.
2. Predictor numbers in ensembles: The number of predictors strongly impacts model performance and can affect disagreement among range loss projections. Our results suggest that optimal performance may be achieved with around 10 predictors, or one predictor per ten presences, if well-designed survey data and diverse predictors are available. For studies using presence-only data and/or exclusively climate predictors, this number may well be lower. The strong dependence of model performance on number of variables makes it straight-forward to optimize this factor for the dataset at hand using block cross-validation.
3. Multicollinearity: In this study, multicollinearity did not strongly affect the performance of model extrapolations, but it distinctly increased projected range loss and the disagreement among range change projections. We recommend keeping absolute Pearson correlation coefficients below 0.7, a boundary recommended elsewhere (Dormann et al., 2013), and one above which consequences in projections became clearly visible.

ACKNOWLEDGEMENTS

This work was supported by the ANR-SNF bilateral project OriginAlps, with grant numbers 310030L_170059 (P.B., N.E.Z., Y. C.) and ANR-16-CE93-004 (W.T.). W.T. also acknowledges the Agence Nationale de la Recherche (ANR) through the "Investissement d'Avenir" grants managed by the ANR (Trajectories: ANR-15-IDEX-02; Montane: OSUG@2020: ANR-10-LAB-56). R.O.W. acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 787638) granted to Catherine Graham. Furthermore, we thank ICP Forest for access to the aggregated tree species data from the Level I network.

DATA AVAILABILITY STATEMENT

The data compiled for the literature analysis are available in Appendix S7. The environmental data used in the main analysis are publicly

available with access information provided in the 'Data' subsection of the Methods. Observational data from ICP Forests can be directly requested online (<http://icp-forests.net/page/data-requests>, accessed September 2019).

ORCID

Philipp Brun  <https://orcid.org/0000-0002-2750-9793>

Wilfried Thuiller  <https://orcid.org/0000-0002-5388-5274>

Rafael O. Wüest  <https://orcid.org/0000-0001-6047-1945>

Niklaus E. Zimmermann  <https://orcid.org/0000-0003-3099-9604>

REFERENCES

- Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, 43, 1223–1232. <https://doi.org/10.1111/j.1365-2664.2006.01214.x>
- Araújo, M. B., & New, M. (2007). Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, 22, 42–47. <https://doi.org/10.1016/j.tree.2006.09.010>
- Beaumont, L. J., Graham, E., Duursma, D. E., Wilson, P. D., Cabrelli, A., Baumgartner, J. B., ... VanDerWal, J. (2016). Which species distribution models are more (or less) likely to project broad-scale, climate-induced shifts in species ranges? *Ecological Modelling*, 342, 135–146. <https://doi.org/10.1016/j.ecolmodel.2016.10.004>
- Bell, D. M., & Schlaepfer, D. R. (2016). On the dangers of model complexity without ecological justification in species distribution modeling. *Ecological Modelling*, 330, 50–59. <https://doi.org/10.1016/j.ecolmodel.2016.03.012>
- Boria, R. A., Olson, L. E., Goodman, S. M., & Anderson, R. P. (2014). Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecological Modelling*, 275, 73–77. <https://doi.org/10.1016/j.ecolmodel.2013.12.012>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.
- Buisson, L., Thuiller, W., Casajus, N., Lek, S., & Grenouillet, G. (2010). Uncertainty in ensemble forecasting of species distribution. *Global Change Biology*, 16, 1145–1157. <https://doi.org/10.1111/j.1365-2486.2009.02000.x>
- Chala, D., Brochmann, C., Psomas, A., Ehrlich, D., Gizaw, A., Masao, C. A., ... Zimmermann, N. E. (2016). Good-bye to tropical alpine plant giants under warmer climates? Loss of range and genetic diversity in *Lobelia rhynchopetalum*. *Ecology and Evolution*, 6, 8931–8941.
- Dormann, C. F., Calabrese, J. M., Guillera-Arroita, G., Matechou, E., Bahn, V., Bartoň, K., ... Hartig, F. (2018). Model averaging in ecology: A review of Bayesian, information-theoretic, and tactical approaches for predictive inference. *Ecological Monographs*, 88, 485–504. <https://doi.org/10.1002/ecm.1309>
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., ... Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36, 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
- Duque-Lazo, J., van Gils, H., Groen, T. A., & Navarro-Cerrillo, R. M. (2016). Transferability of species distribution models: The case of *Phytophthora cinnamomi* in Southwest Spain and Southwest Australia. *Ecological Modelling*, 320, 62–70. <https://doi.org/10.1016/j.ecolmodel.2015.09.019>
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29, 1189–1232. <https://doi.org/10.1214/aos/1013203451>



- García, R. A., Burgess, N. D., Cabeza, M., Rahbek, C., & Araújo, M. B. (2012). Exploring consensus in 21st century projections of climatically suitable areas for African vertebrates. *Global Change Biology*, 18, 1253–1269. <https://doi.org/10.1111/j.1365-2486.2011.02605.x>
- García-Callejas, D., & Araújo, M. B. (2016). The effects of model and data complexity on predictions from species distributions models. *Ecological Modelling*, 326, 4–12. <https://doi.org/10.1016/j.ecolmodel.2015.06.002>
- Greenwell, B., Boehmke, B., Cunningham, J., & Developers, G. B. M. (2018). gbm. Generalized boosted regression models. R package version 2.1.5. Retrieved from <https://CRAN.R-project.org/package=gbm>
- Greg, E. J., Palacios, D. M., Thompson, A., & Chan, K. M. A. (2018). Why less complexity produces better forecasts: An independent data evaluation of kelp habitat models. *Ecography*, 42, 428–443.
- Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135, 147–186. [https://doi.org/10.1016/S0304-3800\(00\)00354-9](https://doi.org/10.1016/S0304-3800(00)00354-9)
- Hao, T., Elith, J., Guiller-Aroita, G., & Lahoz-Monfort, J. J. (2019). A review of evidence about use and performance of species distribution modelling ensembles like BIOMOD. *Diversity and Distributions*, 25(5), 839–852. <https://doi.org/10.1111/ddi.12892>
- Harrell, F. E., Margolis, P. A., Gove, S., Mason, K. E., Mulholland, K. E., Lehmann, D., ... Eichenwald, H. F. (1998). Development of a clinical prediction model for an ordinal outcome. *Statistics in Medicine*, 17, 909–944.
- Hastie, T. (2018). gam. Generalized additive models. R package version 1.16. Retrieved from <https://CRAN.R-project.org/package=gam>
- Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models*. London, UK: Chapman & Hall/CRC.
- Hiederer, R. (2013). Mapping soil properties for Europe – Spatial representation of soil database attributes. EUR25932EN. EUR - Scientific and Technical Research Reports. Scientific and Technical Research Series 47. Publications Office of the European Union. ISBN: 978-92-79-32516-8
- IPCC (2013). *Climate change 2013: The physical science basis*. Cambridge, UK: Cambridge University Press.
- Jones, R. J. A., Hiederer, R., Rusco, E., & Montanarella, L. (2005). Estimating organic carbon in the soils of Europe for policy support. *European Journal of Soil Science*, 56, 655–671. <https://doi.org/10.1111/j.1365-2389.2005.00728.x>
- Karger, D. N., Conrad, O., Böhrer, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., ... Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, 4, 170122. <https://doi.org/10.1038/sdata.2017.122>
- Liaw, A., & Wiener, M. (2002). Classification and regression by random Forest. *R News*, 2, 18–22.
- Lorenz, M. (1995). International co-operative programme on assessment and monitoring of air pollution effects on forests-ICP forests. *Water, Air, & Soil Pollution*, 85, 1221–1226. <https://doi.org/10.1007/BF00477148>
- Marthews, T. R., Dadson, S. J., Lehner, B., Abele, S., & Gedney, N. (2015). High-resolution global topographic index values. NERC Environmental Information Data Centre. <https://doi.org/10.5285/6b0c4358-2bf3-4924-aa8f-793d468b92be>
- McCullagh, P., & Nelder, J. A. (1983). *Generalized linear models*. London, UK: Chapman & Hall.
- Merow, C., Smith, M. J., Edwards, T. C., Guisan, A., McMahon, S. M., Normand, S., ... Elith, J. (2014). What do we gain from simplicity versus complexity in species distribution models? *Ecography*, 37, 1267–1281. <https://doi.org/10.1111/ecog.00845>
- Moreno-Amat, E., Mateo, R. G., Nieto-Lugilde, D., Morueta-Holme, N., Svenning, J.-C.-C., & García-Amorena, I. (2015). Impact of model complexity on cross-temporal transferability in Maxent species distribution models: An assessment using paleobotanical data. *Ecological Modelling*, 312, 308–317. <https://doi.org/10.1016/j.ecolmodel.2015.05.035>
- Petitpierre, B., Broennimann, O., Kueffer, C., Daehler, C., & Guisan, A. (2017). Selecting predictors to maximize the transferability of species distribution models: Lessons from cross-continental plant invasions. *Global Ecology and Biogeography*, 26, 275–287. <https://doi.org/10.1111/geb.12530>
- R Development Core Team (2008). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Randin, C. F., Dirnböck, T., Dullinger, S., Zimmermann, N. E., Zappa, M., & Guisan, A. (2006). Are niche-based species distribution models transferable in space? *Journal of Biogeography*, 33, 1689–1703. <https://doi.org/10.1111/j.1365-2699.2006.01466.x>
- Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guiller-Aroita, G., ... Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40, 913–929. <https://doi.org/10.1111/ecog.02881>
- Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society*, 71, 319–392. <https://doi.org/10.1111/j.1467-9868.2008.00700.x>
- Sanderson, B. M., Knutti, R., & Caldwell, P. (2015). A representative democracy to reduce interdependency in a multimodel ensemble. *Journal of Climate*, 28, 5171–5194. <https://doi.org/10.1175/JCLI-D-14-00362.1>
- Swets, J. A. (1988) Measuring the accuracy of diagnostic systems. *Science (New York, N.Y.)*, 240, 1285–1293.
- Thuiller, W. (2004). Patterns and uncertainties of species' range shifts under climate change. *Global Change Biology*, 10, 2020–2027. <https://doi.org/10.1111/j.1365-2486.2004.00859.x>
- Thuiller, W., Albert, C., Araújo, M. B., Berry, P. M., Cabeza, M., Guisan, A., ... Zimmermann, N. E. (2008). Predicting global change impacts on plant species' distributions: Future challenges. *Perspectives in Plant Ecology, Evolution and Systematics*, 9, 137–152. <https://doi.org/10.1016/j.ppees.2007.09.004>
- Thuiller, W., Guéguen, M., Renaud, J., Karger, D. N., & Zimmermann, N. E. (2019). Uncertainty in ensembles of global biodiversity scenarios. *Nature Communications*, 10, 1446. <https://doi.org/10.1038/s41467-019-09519-w>
- Thuiller, W., Lafourcade, B., Engler, R., & Araújo, M. B. (2009). BIOMOD – A platform for ensemble forecasting of species distributions. *Ecography*, 32, 369–373. <https://doi.org/10.1111/j.1600-0587.2008.05742.x>
- Tóth, B., Weynants, M., Nemes, A., Makó, A., Bilas, G., & Tóth, G. (2015). New generation of hydraulic pedotransfer functions for Europe. *European Journal of Soil Science*, 66, 226–238. <https://doi.org/10.1111/ejss.12192>
- Warren, D. L., Cardillo, M., Rosauer, D. F., & Bolnick, D. I. (2014). Mistaking geography for biology: Inferring processes from species distributions. *Trends in Ecology & Evolution*, 29, 572–580. <https://doi.org/10.1016/j.tree.2014.08.003>
- Wauchope, H. S., Shaw, J. D., Varpe, Ø., Lappo, E. G., Boertmann, D., Lancot, R. B., & Fuller, R. A. (2017). Rapid climate-driven loss of breeding habitat for Arctic migratory birds. *Global Change Biology*, 23, 1085–1094. <https://doi.org/10.1111/gcb.13404>
- Werkowska, W., Márquez, A. L., Real, R., & Acevedo, P. (2017). A practical overview of transferability in species distribution modeling. *Environmental Reviews*, 25, 127–133. <https://doi.org/10.1139/er-2016-0045>
- Zurell, D., Thuiller, W., Pagel, J., Cabral, J. S., Münkemüller, T., Gravel, D., ... Zimmermann, N. E. (2016). Benchmarking novel approaches for modelling species range dynamics. *Global Change Biology*, 22, 2651–2664.



BIOSKETCH

Dr. Philipp Brun is a macroecologist investigating species, trait, and biodiversity distributions with a focus on climate change impact and links to ecosystem functioning. His main study organisms include terrestrial plants and marine plankton. Together with N.E.Z., W.T., and Y.C. he is involved in the 'OriginAlps' project that aims at improving the understanding of historical and contemporary processes that underpin patterns of plant biodiversity in the European Alps.

Author contributions: P.B., N.E.Z., and W.T. conceived the general idea and designed the study with the help of all authors. P.B. performed the analysis and led the writing of the manuscript. All authors interpreted results and significantly contributed to writing and editing.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Brun P, Thuiller W, Chauvier Y, et al.

Model complexity affects species distribution projections under climate change. *J Biogeogr.* 2020;47:130–142.

<https://doi.org/10.1111/jbi.13734>