


REVIEWS AND
SYNTHESES

Species distribution models have limited spatial transferability for invasive species

Chunlong Liu,^{1,2,3,4,*†} 
 Christian Wolter,¹ 
 Weiwei Xian^{4,5}  and
 Jonathan M. Jeschke^{1,2,3} 

Abstract

The reliability of transferring species distribution models (SDMs) to new ranges and future climates has been widely debated. Biological invasions offer the unique opportunity to evaluate model transferability, as distribution data between species' native and introduced ranges are geographically independent of each other. Here, we performed the first global quantitative synthesis of the spatial transferability of SDMs for 235 invasive species and assessed the association of model transferability with the focal invader, model choice and parameterisation. We found that SDMs had limited spatial transferability overall. However, model transferability was higher for terrestrial endotherms, species introduced from or to the Southern Hemisphere, and species introduced more recently. Model transferability was also positively associated with the number of presences for model calibration and evaluation, respectively, but negatively with the number of predictors. These findings highlight the importance of considering the characteristics of the focal invader, environment and modelling in the application and assessment of SDMs.

Keywords

Accuracy, Bayesian inference, biogeographical experiments, biological invasions, ecological niche models, invasive alien species, model extrapolation, model interpolation, model parameterisation, quantitative synthesis.

Ecology Letters (2020) **23**: 1682–1692

INTRODUCTION

Species distribution models (SDMs) have become an essential tool for ecological and evolutionary studies over the last three decades (Guisan and Thuiller, 2005; Araújo *et al.*, 2019). Many researchers have applied SDMs to address historical and phylogeographical questions, to investigate drivers of species distributions, to estimate dynamics of climate niches or to quantify probabilities of disease outbreaks (Guisan *et al.*, 2014; Feng *et al.*, 2019). Given anthropogenic activities and environmental changes of unparalleled magnitude, the role of SDMs in the conservation of biodiversity is paramount (Araújo and Rahbek 2006; Thuiller *et al.*, 2019). In particular, thousands of studies have employed SDMs to forecast species responses to climate changes, to search for new populations of rare species and to predict the impacts of land use on species distributions (e.g. Araújo and Rahbek 2006; Murray *et al.*, 2011; Morán-Ordóñez *et al.*, 2017; Feng *et al.*, 2019). Predictions of species distributions are important for decision making and environmental planning (Wiens *et al.*, 2009; Araújo *et al.*, 2019); however, the reliability of model predictions has been widely debated, leading to increasing

uncertainties in applying SDMs in conservation biogeography (reviewed in Sequeira *et al.*, 2018).

Confidence in transferred models depends on their predictability in new geographical ranges or time periods (i.e. their transferability; Yates *et al.*, 2018). Models can generate accurate predictions within the domain of model calibration (i.e. interpolation), whereas predictive ability might quickly decrease if models are transferred beyond this domain (i.e. extrapolation; Roberts *et al.*, 2017). Recent years have seen mounting efforts to better understand the patterns and determinants of model transferability, but remarkable disputes exist among studies (e.g. Heikkinen *et al.*, 2012; Morán-Ordóñez *et al.*, 2017; Petitpierre *et al.*, 2017). This ongoing debate and conflicting conclusions undermine the confidence in model predictions and pose pressing hurdles on the improvement of model transferability. Recently, 50 experts outlined 12 outstanding challenges related to model transferability which, if addressed, will increase the reliability of SDMs and facilitate their applications (Yates *et al.*, 2018).

True transferability can only be evaluated with data that are spatiotemporally independent of the data used for

¹Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB), Berlin, Germany

²Institute of Biology, Freie Universität Berlin, Berlin, Germany

³Berlin-Brandenburg Institute of Advanced Biodiversity Research (BBIB), Berlin, Germany

⁴CAS Key Laboratory of Marine Ecology and Environmental Sciences, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China

⁵Laboratory for Marine Ecology and Environmental Science, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266071, China

*Correspondence: E-mail: liuchunlong113@gmail.com

†Current Address: Departament de Ciències Ambientals, Facultat de Ciències, Universitat de Girona, Girona, Spain

calibrating a model, because the dependency between data used for calibration and evaluation notoriously leads to underestimating prediction errors (Jeschke and Strayer, 2008; Kharouba *et al.*, 2013; Roberts *et al.*, 2017). Biological invasions represent unique large-scale biogeographical experiments for evaluating model transferability (Jeschke and Strayer, 2008; Petitpierre *et al.*, 2017). Large numbers of invasive species (i.e. non-native species that have established and spread in their new ranges; Blackburn *et al.*, 2011) have been introduced beyond their native ranges, resulting in geographically independent datasets. The pathways and impacts of invasive species have been widely studied (e.g. Saul *et al.*, 2017; Vilà and Hulme, 2017). They are beyond the scope of this study and thus will not be detailed or referenced further. However, they yielded numerous occurrence data in the introduced ranges, and the distributions of invasive species have been systematically investigated across different environmental conditions. SDMs have been widely used to predict the geographical patterns of invasion risks (e.g. Bellard *et al.*, 2013; Hill *et al.*, 2017), in some cases also to evaluate models transferred between species' native and introduced ranges (e.g. Fitzpatrick *et al.*, 2007; Petitpierre *et al.*, 2012).

Model transferability is intrinsically determined by the fitted relationships between environmental predictors and species distributions, which have demonstrated remarkable taxonomic differences (Guisan and Thuiller, 2005). Correspondingly, elucidating variations in model transferability among groups of organisms was identified as the first challenge in Yates *et al.* (2018). Good model transferability has been found in some invasive birds (e.g. Strubbe *et al.*, 2015), insects (e.g. Hill *et al.*, 2017), plants (e.g. Petitpierre *et al.*, 2012) and reptiles (e.g. Tingley *et al.*, 2016), but poor transferability was reported for some invasive crustaceans (e.g. Larson *et al.*, 2010), marine fishes (e.g. Parravicini *et al.*, 2015) and molluscs (e.g. Gallardo *et al.*, 2013). These pervasive discrepancies among taxa are puzzling and further increase uncertainties in the predictions of transferred models (Yates *et al.*, 2018). To date, however, only few studies have compared model transferability across taxonomic groups, and these only included a limited number of taxa (Heikkinen *et al.*, 2012; Morán-Ordóñez *et al.*, 2017).

An additional challenge is to understand how model choice and parameterisation affect transferability (Yates *et al.*, 2018). Identifying the approach with the best predictability has been frequently investigated because different approaches vary in their estimation of species–environment relationships (Qiao *et al.*, 2015; Norberg *et al.*, 2019). Ensemble forecasting (Ensemble) is considered as one of the most promising approaches, as it avoids overreliance on one specific model by averaging predictions from multiple models (Araújo and New, 2007). The maximum entropy method (Maxent; Phillips *et al.*, 2006) is currently a very popular technique for developing SDMs because it only uses presence data and has a convenient interface (Qiao *et al.*, 2015). The type of environmental and distribution data used for calibrating and evaluating models also plays an important role in model transferability (Guisan and Thuiller, 2005). Overparameterising models with excessive predictors can fit complex relationships in the training domain, but very likely yields unreliable predictions in new domains with varying relationships

(Petitpierre *et al.*, 2017; Roberts *et al.*, 2017). Another factor affecting model transferability is the number of presences used in model calibration, as models calibrated with few data points cannot buffer the influence of outliers (Guisan and Thuiller, 2005). Moreover, there is still no commonly agreed upon metric for evaluating model performance (Challenge 12 in Yates *et al.*, 2018; see also Sequeira *et al.*, 2018), which further exacerbates the difficulties of finding a consensus across studies (Breiner *et al.*, 2015; Qiao *et al.*, 2015).

Quantifying and comparing the extent of geographical extrapolation (i.e. beyond the area where training data were collected; Sequeira *et al.*, 2018) versus environmental extrapolation (i.e. beyond the range of predictors for model training and/or beyond the known distribution–predictor relationship; Sequeira *et al.*, 2018) is another challenging task (Yates *et al.*, 2018; Qiao *et al.*, 2019). It is intuitive to assume that models have higher transferability for proximate areas (Roberts *et al.*, 2017), which has been empirically confirmed for some species (e.g. Murray *et al.*, 2011). However, a global review did not find a clear relationship between model transferability and geographic proximity (Yates *et al.*, 2018). Contrasting conclusions also exist for the consequences of environmental extrapolation. Although both simulations (e.g. Qiao *et al.*, 2019) and empirical studies (e.g. Morán-Ordóñez *et al.*, 2017) reported better transferability for areas with similar climate, other studies reported negligible effects of climatic analogy on model transferability (e.g. Petitpierre *et al.*, 2012).

Here, we performed the first global quantitative synthesis of studies that reported the spatial transferability of SDMs for invasive species. Specifically, we focus on two overarching questions: (1) What is the spatial transferability of SDMs in general, and how does it vary across groups of organisms? (2) Is model transferability associated with the focal invader(s) and the model development? To do this, we first standardised evaluation values of different metrics to estimate model transferability and then applied Bayesian hierarchical models to assess the associations of model transferability with 12 factors related to the focal invader(s) and the model development.

MATERIALS AND METHODS

Study compilation

To systematically compile studies that evaluated the spatial transferability of SDMs for invasive species, we applied a four-stage literature search following the PRISMA guidelines (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; Moher *et al.*, 2009; Fig. S1 in Supporting Information; the searches were conducted in January 2019). First, we searched the literature in Clarivate Analytics' Web of Science database (WoS) with the following terms: “(introduce* OR inva* OR non-native OR nonnative OR exotic OR naturaliz* OR nonindigenous OR non-indigenous) AND (transferability OR (niche AND climat*))”. We included “niche” and “climat*” in the search query because evaluating model transferability is one of two main approaches to assess the dynamics of a species' climate niche (Guisan *et al.*, 2014). Second, we screened the titles and abstracts of the publications returned by the search, and we only kept studies that quantitatively

assessed the predictive ability of SDMs. Third, we screened the reference sections of these publications to find further relevant publications. In total, the titles and abstracts of 1906 publications were screened for identifying potentially relevant studies. Fourth, we consulted the full texts of 256 potentially relevant studies and selected those studies that assessed geographically transferred SDMs. We excluded studies that only predicted the distribution of invasive species without quantitatively assessing model transferability. The final dataset consists of 65 studies encompassing 235 unique invasive species; it is provided as Table S1.

Data compilation

For each study, we first categorised each model according to the geographical range where presence data were collected for model training: native-range, introduced-range or global (i.e. both native and introduced ranges) models. Values of different metrics for evaluating the performance of models to predict native, introduced and global occurrences (i.e. presences and/or absences) were then compiled. We conducted the extensive collection of available information for factors related to focal invader and model parameterisation (Table 1). Species were classified into four groups of organisms: Plants, Terrestrial endotherms (birds and mammals), Terrestrial ectotherms (reptiles, insects and amphibians) and Aquatic species (crustaceans, corals, algae, molluscs and fishes). The native and introduced continents for each species followed the authors' definition, and continents were considered as the same if a species was introduced within one continent. The number of species' native or introduced continent(s) was classified into (1) only one continent and (2) more than one continent, because model transferability is strongly influenced by the number of

biogeographical regions (Morán-Ordóñez *et al.*, 2017). For species with only one native or introduced continent, the continent was coded as the Northern Hemisphere (Asia, Europe and North America) or the Southern Hemisphere (Africa, Australia and South America) to account for the much higher uniqueness of species from the Southern Hemisphere (Holt *et al.*, 2013). The earliest year of recorded introduction in the introduced continent(s) was collected from each study and supplemented with data from the Global Alien Species First Records Database (Seebens *et al.*, 2018; accessed in March 2019). Introduction intentionality of each species followed the classification in Saul *et al.* (2017): (1) Either intentional or unintentional, and (2) Both intentional and unintentional. Data on parameters associated with the modelling were collected for factors that were available in all studies, including the spatial resolution of both predictor and presence data, the number of predictors, and the number of presences in the native (native presences) and introduced ranges (introduced presences). The developed models were classified as (1) Ensemble, (2) only Maxent (Maxent) and (3) other techniques (e.g. Random forests, Generalised linear models and Artificial neural networks; see Table S1 for 13 other techniques). Ensemble and Maxent were categorised as separate groups because they are fundamentally different from other techniques/approaches (see above) and have been widely used (e.g. Yackulic *et al.*, 2013; Feng *et al.*, 2019; see also Results). Given the various ways of conducting background sampling (e.g. VanDerWal *et al.*, 2009; Feng *et al.*, 2019), we only set a binary variable to represent whether pseudo-absence data were selected from constrained background areas. The analogy of environments between calibration and evaluation domains is also identified as an important factor for model transferability (Strubbe *et al.*, 2015; Petitpierre *et al.*, 2017), and we followed the

Table 1 Description and summary data for factors related to focal invader and model development. For the seven categorical factors, the number of models is summarised for each group. For the five consecutive factors, the range, mean and coefficient of variation (CV) of values are summarised

Factors		Descriptions and summary data
Categorical	Group of organisms	Species were classified into Plants ($N = 163$), Terrestrial endotherms ($N = 150$), Terrestrial ectotherms ($N = 76$) and Aquatic species ($N = 52$).
	Native range	The range from which a species was introduced, which we first classified into More than one continent ($N = 70$) and Only one continent. Only one continent was further divided into the Northern Hemisphere ($N = 282$) or Southern Hemisphere ($N = 89$).
	Introduced range	The range to which a species was introduced, following the classification of the native range (More than one continent, $N = 47$; Northern Hemisphere, $N = 307$; Southern Hemisphere, $N = 87$).
	Introduction intentionality	The intentionality of species introductions was classified into Either intentional or unintentional ($N = 359$), and Both intentional and unintentional ($N = 82$).
	Model group	Techniques used for developing models were classified into Ensemble forecasting ($N = 151$), only Maxent ($N = 123$) and other techniques ($N = 167$).
	Constrained absence	A binary variable to represent whether pseudo-absence data were selected from constrained background areas: Constrained ($N = 242$) and Unconstrained ($N = 199$).
	Analogous environment	A binary variable to represent whether SDMs were only transferred within analogous environments between ranges: Analogous ($N = 229$) and Non-analogous ($N = 212$).
Consecutive	Year of introduction	The earliest year in which an invasive species was reported to have been introduced to the introduced range (Range: 500–2009; Mean: 1855.21; CV: 0.1).
	Resolution (°)	The spatial resolution of both predictor and presence data (Range: 0.000278–0.83; Mean: 0.37; CV: 0.64).
	Number of predictors	The number of predictors used for developing models (Range: 2–30; Mean: 8.58; CV: 0.39).
	Number of native presences	The number of presences compiled in the native range (Range 11–24599; Mean: 1401.61; CV: 1.71).
	Number of introduced presences	The number of presences compiled in the introduced range (Range: 5–6079; Mean: 414.5; CV: 1.94).

authors' conclusion on whether models were only transferred within analogous environments (i.e. conditions present in both native and introduced ranges; Petitpierre *et al.*, 2012). We requested the data of evaluation values and factors from the authors if the data were not available in the publications. Of the ten authors contacted, six responded and shared their data. The studies of those four not responding authors were excluded from our analyses.

A major hurdle for synthesising findings of model transferability across studies is the substantial heterogeneity in metrics used for evaluating model performance (Table S1 and S2). Here, we standardised evaluation values of all metrics to the same scale from 0 to 10 as follows: $s = (o-w)/(b-w) \times 10$, where s is the standardised value, o is the original value, and w and b are the values representing the worst and best model performance of the metric, respectively (Table S2). The proportional standardisation makes evaluation values of different metrics comparable: 0 represents the worst performance, whereas 10 represents the best performance. Taking the area under the ROC curve (AUC; Fielding and Bell, 1997) and Boyce Index (Boyce; Boyce *et al.*, 2002) as examples, AUC = 0.8 and Boyce = 0.6 would be both standardised to $s = 8$ because AUC ranges from 0 to 1 and Boyce from -1 to 1. We excluded studies with an evaluation index that could not be standardised from 0 to 10 (e.g. Transferability index by Duque-Lazo *et al.*, 2016). To avoid pseudo-replication of model evaluation, we averaged the standardised values of different metrics for each single model prediction. This is because evaluation values of different metrics have been reported to be highly correlated (e.g. Breiner *et al.*, 2015; Norberg *et al.*, 2019), and differences in evaluation values between two metrics were < 1 for most model predictions (50.7%) in our study (Fig. S2).

Defining accuracy and transferability

We explicitly defined model transferability as the performance of the native-range model that was extrapolated to correctly predict occurrences in the introduced range. Indeed, models were more frequently calibrated in the native range and transferred to introduced ranges (441 models from 63 studies) than vice versa (72 models from 21 studies). Model accuracy was defined as the performance of the native-range model that was interpolated to correctly predict occurrences in the native range (Sequeira *et al.*, 2018; Araújo *et al.*, 2019).

Estimation of accuracy and transferability

The overall accuracy and transferability across all species and for each group of organisms were estimated by Bayesian inference techniques using the 'BEST' package (Kruschke and Meredith, 2018) in R software (v 3.5.0.) (R Development Core Team 2018). Bayesian methods provide a robust estimation of parameters and the degree of uncertainty by controlling biases caused by small sample sizes and potential outliers (Lemoine, 2019). We adopted weakly informative priors by setting the average of input values as mean and 2 as standard deviation

(s.d.) (Lemoine, 2019). We ran four chains of 5000 iterations with the first 2000 iterations as the burn-in and kept 12 000 posterior values. The Highest Density Interval (HDI) of retained posterior values was used for describing credible intervals of estimates. The semi-quantitative scheme suggested by Thuiller *et al.* (2005) was used in a slightly modified way for ranking accuracy and transferability: Excellent (9–10), Good (8–9), Fair (7–8), Poor (6–7), Fail (5–6) and Worse than random (< 5).

Relationship between accuracy and transferability

The relationship between accuracy and transferability for all species and for each group of organisms was fitted by a Bayesian multilevel model using the 'brms' package (Bürkner, 2017). We set the study as the random variable to account for the heterogeneity of focal invader and model development among studies. Weakly informative priors were adopted for slope (mean = 0; s.d. = 2) and intercept (mean = 0; s.d. = 10). Good convergence of chains was confirmed by values of Rhat (the potential scale-reduction factor) below 1.01 for all models (Bürkner, 2017).

Bayesian hierarchical analysis

We assessed the associations of model transferability and accuracy with 12 covariates related to focal invader and model development for all species and for each group of organisms by Bayesian hierarchical models using the 'brms' package (Bürkner, 2017). Bayesian hierarchical models allow us to specifically assess the effect of one covariate after controlling for influences of other covariates. To facilitate the comparison in effect sizes of covariates, we standardised each continuous covariate with the mean of 0 and the standard deviation of 2. Finally, we compared the performance of global, native-range and introduced-range models by Bayesian hierarchical models, respectively, to correctly predict native and introduced occurrences. For each of the six model pairs, we set model performance as the response variable and the compared models as the categorical variable. We included 12 covariates to control for their influence on model performance. The comparisons were also conducted between accuracy and transferability, and between reciprocal predictions (i.e. native-range model used to predict introduced occurrences, and introduced-range model to predict native occurrences; Fitzpatrick *et al.*, 2007). For each Bayesian hierarchical model, we adopted weakly informative priors for all variables (mean = 0; s.d. = 2) and for the global intercept (mean = 0; s.d. = 10). Estimates of the mean and credible interval of each covariate were based on 12 000 posterior values generated from four chains of 5000 iterations with the first 2 000 iterations as the burn-in. The study was set as the random variable to control for heterogeneous covariates among studies. The goodness of fit was evaluated by Bayesian R^2 values that represent the variance of the predicted values divided by the sum of the variance of predicted values and the expected variance of the errors (Gelman *et al.*, 2019). Good convergence was verified by values of Rhat below 1.01 for all models (Bürkner, 2017).

RESULTS

Study composition

Studies that evaluated model transferability mainly focused on terrestrial invasive species (83.6%), with only nine studies for aquatic species (Fig. S3). Large differences between terrestrial and aquatic systems were also found in the numbers of models (88.2% and 11.8%, respectively), with the greatest number of studies (37.3%) and models (37.0%) for plants (Table 1; Fig. S3). Additionally, most studies (73.8%) evaluated model transferability for only one species, and no study focused on microorganisms (Table S1). Differences in popularity of techniques/approaches were also substantial: almost half of studies (49.2%) applied only Maxent to develop SDMs, although the number of models was rather close among Ensemble, Maxent and other techniques (Table 1; Fig. S3).

Model transferability among continents

The evaluation of model transferability showed great variations among continents. Most models were developed for

species that were introduced from a continent of the Northern Hemisphere (63.3%) and for species introduced to a continent with high level of economic development (Australia, Europe and North America; 82.3%) (Fig. 1). Higher transferability was found for species that were native in more than one continent and introduced to North America (7.95 ± 1.55 ; hereafter, mean transferability \pm 95% confidence intervals), and for species introduced to Australia (7.87 ± 0.33). By contrast, model transferability was lower for species in Europe that were introduced from Asia (4.87 ± 0.78) and from North America (5.59 ± 0.85).

Estimation of accuracy and transferability

The overall model accuracy (9.39; mean effect size represented by the mean of Bayesian posterior values) was much higher than model transferability (6.64) (Fig. 2), suggesting that models performed excellently for interpolating, but poorly when being extrapolated to new areas. A higher accuracy than transferability was found in all groups of organisms, although there were interesting variations among groups: model

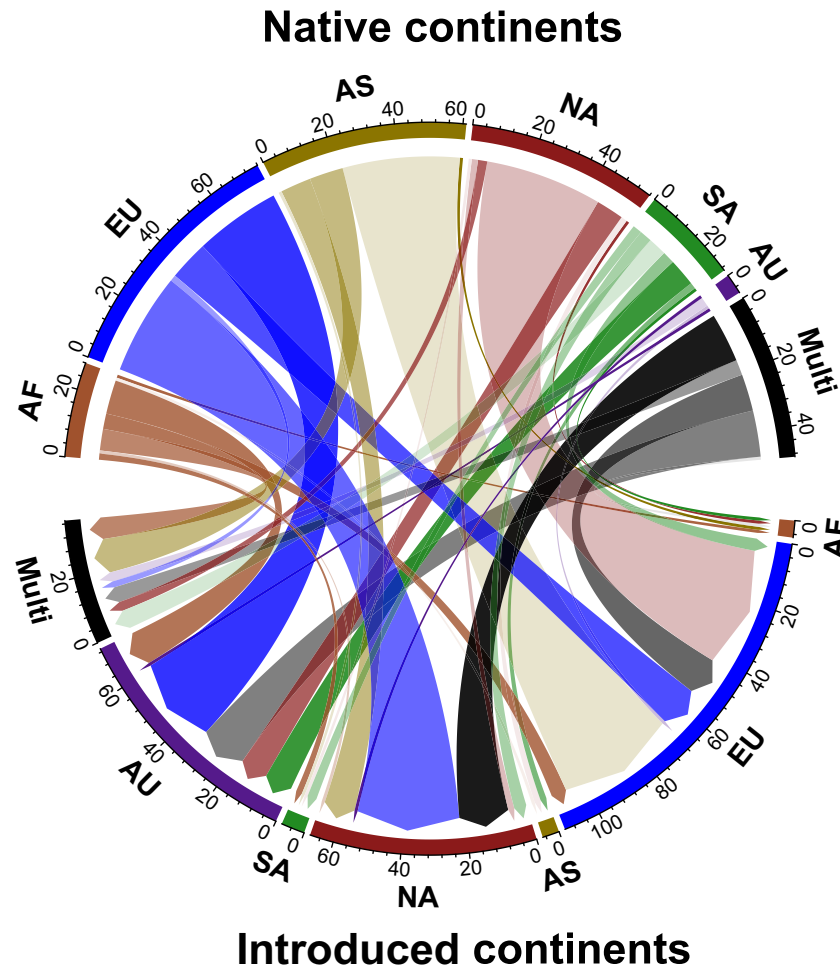


Figure 1 Model transferability among continents. Each continent is represented by a segment with a specific colour. AF, Africa; AS, Asia; AU, Australia; EU, Europe; NA, North America; SA, South America; Multi, More than one continent. The width of the arrow represents the number of species that were introduced from the native continents (upper half) to the introduced continents (lower half). The transparency of the arrow represents the averaged transferability across all models between corresponding native and introduced continents, with darker colours indicating higher transferability and lighter colours indicating lower transferability.

accuracy was excellent for terrestrial endotherms (9.9), plants (9.37) and terrestrial ectotherms (9.33), but only fair for aquatic species (7.89). The highest transferability was found for terrestrial endotherms (7.33) and the lowest for aquatic species (4.79). We did not detect particular changes in accuracy or transferability over the past 14 years, with models consistently exhibiting excellent accuracy but only poor or even failed transferability (Fig. S4).

There were considerable differences in the performances of global, native-range and introduced-range models to predict native and introduced occurrences (Fig. S5). In both native and introduced ranges, models trained there clearly outperformed models transferred from the other range, highlighting the risks associated with model extrapolation. By contrast, predictive performances were very close for models reciprocally transferred between native and introduced ranges. The comparison between accuracy and transferability confirmed the decreased performance of native-range models to predict introduced occurrences. Notably, performances of global models in native and introduced ranges were almost equivalent to those of the native-range and introduced-range models respectively.

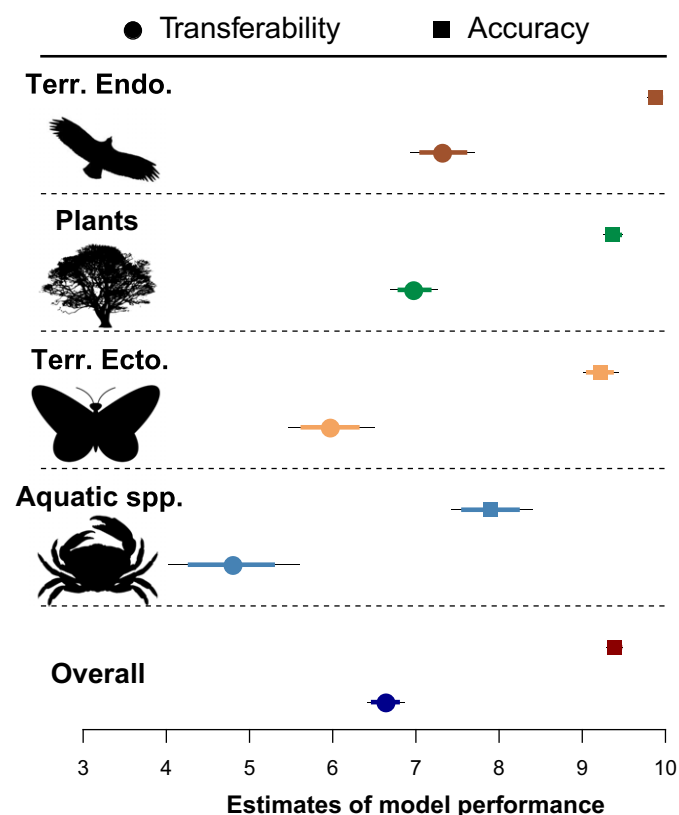


Figure 2 Estimated model accuracy and transferability for all species (Overall) and for each group of organisms (Terrestrial endotherms, Plants, Terrestrial ectotherms and Aquatic species). Means of 12 000 Bayesian posterior values are shown from four chains with 80% Highest Density Interval (HDI; coloured lines) and 95% HDI (thin lines). Model accuracy and transferability were classified into: Excellent (9–10), Good (8–9), Fair (7–8), Poor (6–7), Fail (5–6) and Worse than random (< 5).

Relationship between accuracy and transferability

Overall, there was a positive correlation between model accuracy and transferability across all species [$\beta = 0.34$ (0.05–0.63); mean of posterior distribution and 95% credible interval] (Fig. 3), indicating that models with better performance for interpolation were more likely to correctly predict species distributions in new ranges. This positive correlation between accuracy and transferability was also found in plants [$\beta = 0.69$ (0.09–1.26)], while the other three groups showed no clear relationship (credible interval overlapped with zero).

Influence of factors

Model transferability was closely associated with all characters of focal invaders considered here (Fig. 4; Fig. S6). Variations in transferability among groups of organisms were in agreement with patterns separately estimated for each group, with the highest transferability for terrestrial endotherms. Transferability was also higher if models were developed for species that were introduced from/to the Southern Hemisphere or introduced to more than one continent, whereas it was lower for species with more than one native continent. Moreover, higher transferability was found for species that were introduced more recently or introduced via both intentional and unintentional pathways.

Model group and parameterisation also affected transferability, although differences in effects of relevant factors were rather small (Fig. 4; Fig. S6). We found no evidence of improved transferability for Ensemble and Maxent when compared with other techniques. The number of predictors was negatively related to model transferability, suggesting the deteriorating effects of increased model complexity. By contrast, there were positive correlations of transferability with the number of native and introduced presences, respectively. Additionally, we found no clear influence of environmental analogy on model transferability.

Associations of model transferability with covariates were different among four groups of organisms (Fig. 4; Fig. S7–S10). The magnitude of association was much stronger for plants and weaker for aquatic species, and even contrasting patterns were found between groups. For example, the association of the year of introduction with transferability was negative for terrestrial ectotherms, but positive for plants and terrestrial endotherms. Selecting pseudo-absences from the constrained background impaired model transferability for plants, but improved transferability for terrestrial ectotherms. In addition, model transferability clearly decreased for terrestrial ectotherms at finer resolutions.

None of the covariates related to the focal invader was associated with model accuracy (Fig. S11), revealing the negligible role of species characters on model interpolation. Instead, model accuracy was positively correlated with resolution and the number of native presences, suggesting a higher accuracy of models developed at a finer resolution and with more presences. We also found a slightly negative influence of Maxent for terrestrial ectotherms. For plant species, a lower accuracy was found for models that were calibrated with more predictors or fewer native presences.

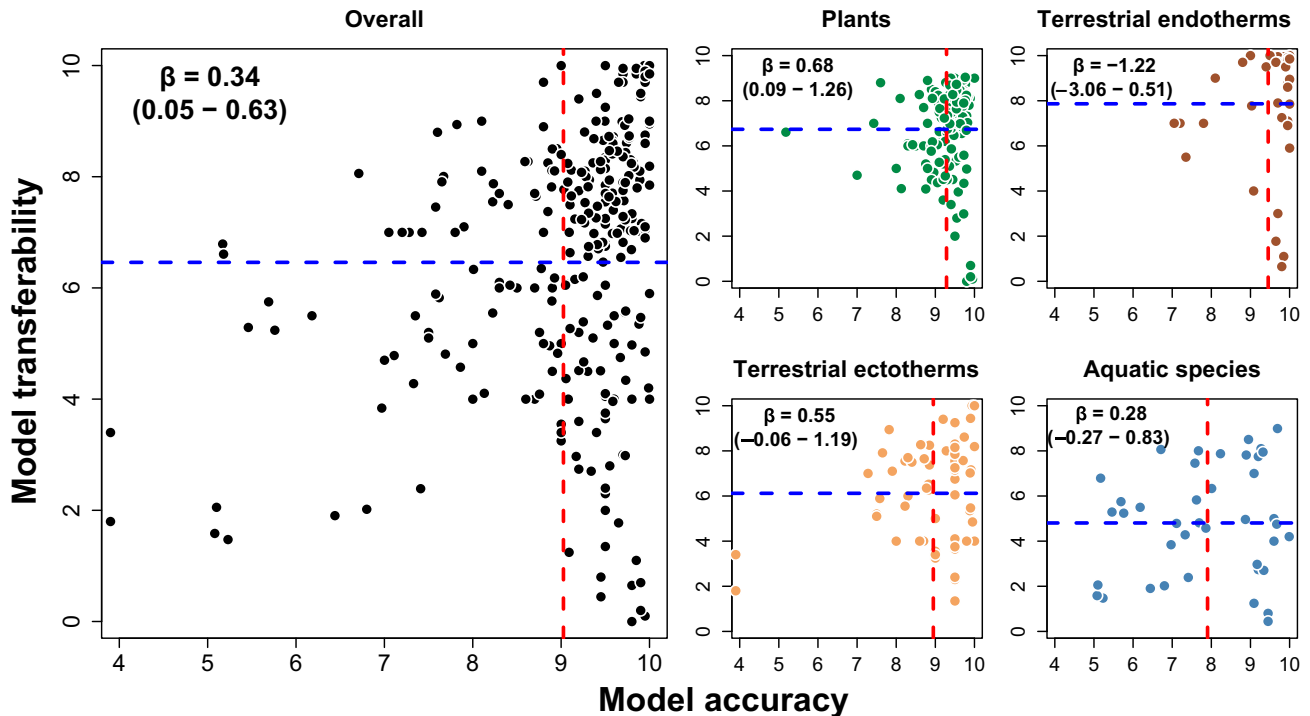


Figure 3 Relationships between model accuracy and transferability. The relationship was separately fitted for all species (Overall) and for each group of organisms by Bayesian multilevel model with study as the random variable. The mean (β) and 95% credible interval of 12 000 posterior values estimated from four chains are shown in each panel. Red and blue dotted lines indicate the mean of accuracy and transferability, respectively, that were averaged across original values of corresponding models.

DISCUSSION

The example of invasive species with geographically independent distributions in their native and introduced ranges revealed a generally limited spatial transferability of SDMs to new areas. Most previous studies assessed the performance of transferred models solely based on the evaluation of model accuracy by means of cross-validation, i.e. the evaluation dataset is a subset of the same data used for model calibration (Roberts *et al.*, 2017). However, our results suggest that evaluating models according to model accuracy will be overly optimistic because calibration and evaluation data are not really independent (Roberts *et al.*, 2017). If covariates in calibration and evaluation data are structured in the same way, the predictive ability is evaluated for model interpolation rather than extrapolation, leading to the selection of highly parameterised models (Dormann *et al.*, 2008; Petitpierre *et al.*, 2017). The underestimation of prediction error by cross-validation could also explain the excellent performance of global models in both native and introduced ranges because training and evaluating sets of global models are inevitably parts of the global distribution data.

Remarkable variations in model transferability among groups of organisms indicate the importance of considering the taxonomic group and type of environment when developing models. The higher transferability for terrestrial endotherms is consistent with previous findings (e.g. Heikkinen *et al.*, 2012; Morán-Ordóñez *et al.*, 2017). Compared to other taxa, distributions of endotherms are more constrained

by broad-scale climatic factors, which are the only modelling predictors in most studies (Dormann *et al.*, 2008; Heikkinen *et al.*, 2012). Instead, smaller organisms tend to have higher plasticity and shorter generation times, leading to faster adaptation to new climatic conditions; and a changed climate–distribution relationship results in reduced model transferability (Soininen *et al.*, 2013; Soininen and Luoto, 2014). For aquatic species, the low model transferability might be directly associated with less comprehensive distribution data because species are more difficult to be directly observed and recorded under water. Small sample sizes and unknown data quality (e.g. because of unreported bias) can increase uncertainty of predictions and consequently hamper model transferability (Dormann *et al.*, 2008; Yackulic *et al.*, 2013). Moreover, SDMs for species native to more than one continent show much lower transferability, indicating a certain difficulty to predict the distribution of widespread species (Qiao *et al.*, 2015; Morán-Ordóñez *et al.*, 2017). Models developed for species with large geographic range suffer higher uncertainty resulting from more heterogeneous calibration data, a lower prevalence of presences, limited or biased sampling coverage (e.g. Dormann *et al.*, 2008; Morán-Ordóñez *et al.*, 2017). An effective strategy to improve model predictability is developing species-specific models or models for functional groups (Guisan and Thuiller, 2005).

Close associations of model transferability with the year and pathway of introduction emphasise the need for integrating these factors into model development and assessment (Bellard *et al.*, 2013; Donaldson *et al.*, 2014). Higher

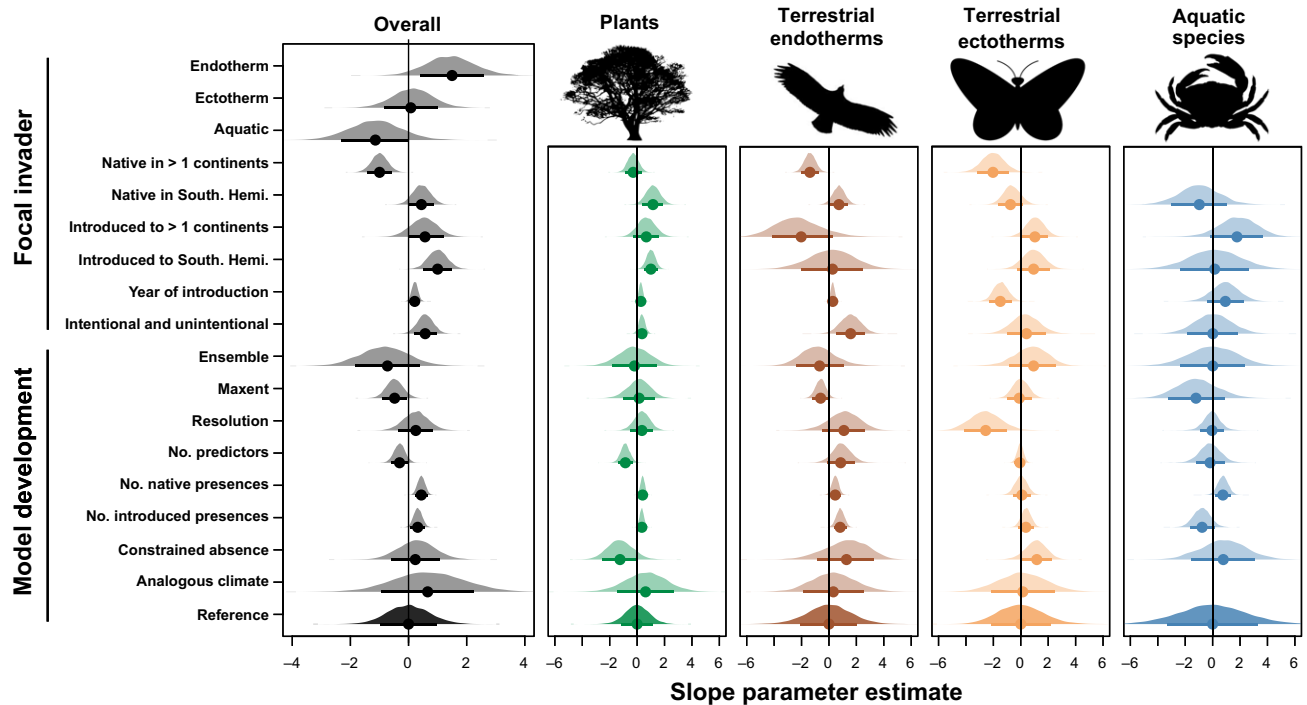


Figure 4 Relationships between model transferability and 12 covariates. The relationship was assessed for all species (Overall; $R^2 = 0.45$) and for Plants ($R^2 = 0.62$), Terrestrial endotherms ($R^2 = 0.32$), Terrestrial ectotherms ($R^2 = 0.45$) and Aquatic species ($R^2 = 0.39$) respectively. The reference categories selected for assessing relative effects of categorical variables are as follows: Plants, Native in the Northern Hemisphere, Introduced to the Northern Hemisphere, Introduced either intentionally or unintentionally, Other techniques, Unconstrained absence and Non-analogous environment (see also Fig. S6-S10). Each posterior probability distribution represents the mean (symbol centre) and 80% credible interval (thick line) of 12 000 Bayesian posterior values estimated from four chains. The vertical line indicates the mean of posterior values for the reference category. A positive value indicates higher model transferability compared to the reference line category for categorical variables, or a positive relationship with model transferability for continuous variables

transferability for species with shorter residence time challenges the conclusions of empirical studies that distributions of invasive species could be more accurately predicted by models developed in later stages of invasion (e.g. Václavík and Meentemeyer, 2012), because species introduced more recently are still spreading and have not yet reached equilibrium distributions (Tingley *et al.*, 2016). However, species with longer residence time are also more likely to spread to areas with environmental conditions very different from their native ranges. If models are extrapolated beyond the environmental conditions of the training domain, model transferability can be severely impaired (Petitpierre *et al.*, 2012; Qiao *et al.*, 2019). The higher transferability for species that were introduced both intentionally and unintentionally reflects the role of propagule pressure (i.e. the number of individuals introduced; Blackburn *et al.*, 2011) on predicting distributions of invasive species. Species introduced via more pathways have a higher chance to overcome biogeographic barriers and arrive in sites with suitable environmental conditions (Donaldson *et al.*, 2014; Saul *et al.*, 2017), contributing to the improvement in model transferability.

Our study addresses the important influence of model choice and parameterisation on model transferability (Guisan and Thuiller, 2005; Wisz *et al.*, 2008). The lower transferability of models developed with a higher number of predictors confirms the deteriorating effects of over-parameterisation on model extrapolation (Petitpierre *et al.*, 2017). Highly

parameterised models can fit well to complex interactions within the training domain, but the confidence in predictions might largely decrease when models are extrapolated to other domains with different data structures (Roberts *et al.*, 2017; Qiao *et al.*, 2019). Reducing model complexity by only including the most influential predictors on species distributions is therefore an appropriate way to improve model transferability (Norberg *et al.*, 2019). The number of presences used for calibrating and evaluating models is also crucial because sufficient sample sizes allow models to more accurately capture species–environment relationships over species’ ranges and reinforce model transferability across regions. Surprisingly, we found a slightly lower transferability of models developed with Maxent, which has been regarded as the gold standard for SDMs (Qiao *et al.*, 2015). Lower transferability for Maxent can be partly attributed to its “clamping” strategy for extrapolating: by default, Maxent first fixes a relative suitability value at the extreme (maximum or minimum) of training dataset, and then assigns the value to all more extreme points beyond the training dataset (Feng *et al.*, 2019), without considering species response curves to changing environments.

Despite offering new insights into understanding model transferability, our study has several limitations. One key assumption underlying model transfer is that species retain a stable relationship with environments in space and time (“Niche conservatism”; Petitpierre *et al.*, 2012). Several studies included here have reported higher model transferability for

invasive species with a more stable niche between native and introduced ranges (e.g. Petitpierre *et al.*, 2012; Strubbe *et al.*, 2015). However, the assumption of niche conservatism has been frequently violated in invasive species that change the realised niche during invasions (Guisan *et al.*, 2014). Niche changes in invasive species can be attributed to genetic adaptations to environmental conditions in the introduced range or the mixing of native populations with different genetic structures (Fitzpatrick *et al.*, 2007; Hill *et al.*, 2017). More importantly, the species' realised niche in the new range is strongly shaped by biotic interactions and dispersal limitations (Pearman *et al.*, 2008; Hill *et al.*, 2017). In addition, source populations of invasive species probably only represent a small part of the fundamental niche if divergent phylogenetic lineages exist in species' native range (Jeschke and Strayer, 2008; Tingley *et al.*, 2016). Changes in the realised niche misspecify the estimated species–environment relationship and in turn impede the forecasting of species distributions. While niche change is considered to be an important factor in model development and evaluation (Pearman *et al.*, 2008), no study has synthesised its effect on model transferability yet. Another source of uncertainty is the dissimilarity of environmental conditions between regions. If models are transferred between regions with dissimilar environments, the loss of confidence in model predictions can be tremendous (Roberts *et al.*, 2017; Qiao *et al.*, 2019). At last, model transferability is driven by other determinants, such as the prevalence of species (e.g. Guisan and Thuiller, 2005) or the number and extent of background data in some modelling approaches (e.g. VanDerWal *et al.*, 2009; Barbet-Massin *et al.*, 2012). However, corresponding data are not available in most SDM studies (Feng *et al.*, 2019), precluding our ability to assess the roles of those factors. We suggest future studies to report necessary modelling information for facilitating a better understanding on the mechanisms underlying model transferability (Sequeira *et al.*, 2018; Feng *et al.*, 2019).

Biological invasions have caused massive alterations of biodiversity patterns in the Anthropocene (Blackburn *et al.*, 2011; Vilà and Hulme, 2017). Forecasting the potential distribution of invasive species remains a core goal in conservation biology (Jeschke and Strayer, 2008). SDMs continue to represent one of the most promising tools (Bellard *et al.*, 2013; Hill *et al.*, 2017); however, our study underlines the challenges of applying SDMs to predict species distributions in new ranges and highlights the need for considering the characteristics of the species, environment and modelling approach. According to our findings, an appropriate way to improve model transferability could be developing models with increased number of presences or to decrease the number of predictors. Given that a 'silver bullet' model with the best performance under all circumstances is impossible to develop (Qiao *et al.*, 2015; Yates *et al.*, 2018), scientists and stakeholders should balance realism, accuracy and generality when predicting invasion risks at large spatial scale.

ACKNOWLEDGEMENTS

CL was sponsored by the International IGB Fellowship Program in Freshwater Science, the Second Tibetan Plateau Scientific Expedition and Research Program (STEP) (no.

2019QZKK0501), the National Natural Science Foundation of China (no. 31472016) and AlienScenarios Project PCI2018-092966 funded by FEDER/Ministerio de Ciencia e Innovación – Agencia Estatal de Investigación. MJM was supported by a Heisenberg grant from the Deutsche Forschungsgemeinschaft (DFG; JE 288/9-2). WX was supported by the National Natural Science Foundation of China (no. 31872568) and the Natural Science Foundation of China–Shandong Joint Fund for Marine Ecology and Environmental Sciences (no. U1606404). We are grateful to Achyut Banerjee, Blaise Petitpierre, Diederik Strubbe, Ileana Herrera, Maxwell Obiakara and Wanwan Liang for providing data, and to Phil Bouchet for comments on an earlier manuscript version. Comments from two anonymous reviewers also greatly improved the manuscript. Open access funding enabled and organized by Projekt DEAL.

AUTHORSHIP

CL and MJM conceived the idea and designed the study. CL ran the analyses and led the writing, with substantial input from MJM and CW. All co-authors commented on and approved the manuscript.

PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1111/ele.13577>.

DATA AVAILABILITY STATEMENT

All data supporting the results are available in Table S1. All R code is available from the Figshare Repository: 10.6084/m9.figshare.12612785.

REFERENCES

- Araújo, M.B., Anderson, R.P., Márcia Barbosa, A., Beale, C.M., Dormann, C.F., Early, R. *et al.* (2019). Standards for distribution models in biodiversity assessments. *Sci. Adv.*, 5, eaat4858.
- Araújo, M.B. & New, M. (2007). Ensemble forecasting of species distributions. *Trends Ecol. Evol.*, 22, 42–47.
- Araújo, M.B. & Rahbek, C. (2006). How does climate change affect biodiversity? *Science*, 313, 1396–1397.
- Barbet-Massin, M., Jiguet, F., Albert, C.H., Thuiller, W., Barbet-Massin, M., Jiguet, F. *et al.* (2012). Selecting pseudo-absences for species distribution models: how, where and how many? *Methods Ecol. Evol.*, 3, 327–338.
- Bellard, C., Thuiller, W., Leroy, B., Genovesi, P., Bakkenes, M. & Courchamp, F. (2013). Will climate change promote future invasions? *Glob. Chang. Biol.*, 19, 3740–3748.
- Blackburn, T.M., Pyšek, P., Bacher, S., Carlton, J.T., Duncan, R.P., Jarošík, V. *et al.* (2011). A proposed unified framework for biological invasions. *Trends Ecol. Evol.*, 26, 333–339.
- Boyce, M.S., Vernier, P.R., Nielsen, S.E. & Schmiegelow, F.K.A. (2002). Evaluating resource selection functions. *Ecol. Modell.*, 157, 281–300.
- Breiner, F.T., Guisan, A., Bergamini, A. & Nobis, M.P. (2015). Overcoming limitations of modelling rare species by using ensembles of small models. *Methods Ecol. Evol.*, 6, 1210–1218.
- Bürkner, P.C. (2017). brms: an R package for Bayesian multilevel models using Stan. *J. Stat. Softw.*, 80, 1–28.
- Donaldson, J.E., Hui, C., Richardson, D.M., Robertson, M.P., Webber, B.L. & Wilson, J.R.U. (2014). Invasion trajectory of alien trees: the role of introduction pathway and planting history. *Glob. Chang. Biol.*, 20, 1527–1537.

- Dormann, C.F., Purschke, O., Márquez, J.R.G., Lautenbach, S. & Schröder, B. (2008). Components of uncertainty in species distribution analysis: a case study of the Great Grey Shrike. *Ecology*, 89, 3371–3386.
- 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. *Ecol. Modell.*, 320, 62–70.
- Feng, X., Park, D.S., Walker, C., Peterson, A.T., Merow, C. & Papeş, M. (2019). A checklist for maximizing reproducibility of ecological niche models. *Nat. Ecol. Evol.*, 3, 1382–1395.
- Fielding, A.H. & Bell, J.F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.*, 24, 38–49.
- Fitzpatrick, M.C., Weltzin, J.F., Sanders, N.J. & Dunn, R.R. (2007). The biogeography of prediction error: why does the introduced range of the fire ant over-predict its native range? *Glob. Ecol. Biogeogr.*, 16, 24–33.
- Gallardo, B., zu Ermgassen, P.S.E. & Aldridge, D.C. (2013). Invasion ratcheting in the zebra mussel (*Dreissena polymorpha*) and the ability of native and invaded ranges to predict its global distribution. *J. Biogeogr.*, 40, 2274–2284.
- Gelman, A., Goodrich, B., Gabry, J. & Vehtari, A. (2019). R-squared for Bayesian regression models. *Am. Stat.*, 73, 307–309.
- Guisan, A., Petitpierre, B., Broennimann, O., Daehler, C. & Kueffer, C. (2014). Unifying niche shift studies: insights from biological invasions. *Trends Ecol. Evol.*, 29, 260–269.
- Guisan, A. & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.*, 8, 993–1009.
- Heikkinen, R.K., Marmion, M. & Luoto, M. (2012). Does the interpolation accuracy of species distribution models come at the expense of transferability? *Ecography*, 35, 276–288.
- Hill, M.P., Gallardo, B. & Terblanche, J.S. (2017). A global assessment of climatic niche shifts and human influence in insect invasions. *Glob. Ecol. Biogeogr.*, 26, 679–689.
- Holt, B.G., Lessard, J.P., Borregaard, M.K., Fritz, S.A., Araújo, M.B., Dimitrov, D. *et al.* (2013). An update of Wallace's zoogeographic regions of the world. *Science*, 339, 74–78.
- Jeschke, J.M. & Strayer, D.L. (2008). Usefulness of bioclimatic models for studying climate change and invasive species. *Ann. N.Y. Acad. Sci.*, 1134, 1–24.
- Kharouba, H.M., Mccune, J.L., Thuiller, W. & Huntley, B. (2013). Do ecological differences between taxonomic groups influence the relationship between species' distributions and climate? A global meta-analysis using species distribution models. *Ecography*, 36, 657–664.
- Kruschke, J. & Meredith, M. (2018). BEST: Bayesian estimation supersedes the t-test. R package version 0.5.1. <https://CRAN.R-project.org/package=BEST>
- Larson, E.R., Olden, J.D. & Usio, N. (2010). Decoupled conservatism of Grinnellian and Eltonian niches in an invasive arthropod. *Ecosphere*, 1, 1–13.
- Lemoine, N.P. (2019). Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. *Oikos*, 128, 912–928.
- Moher, D., Liberati, A., Tetzlaff, J. & Altman, D.G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. *PLoS Medicine*, 6, e1000097.
- Morán-Ordóñez, A., Lahoz-Monfort, J.J., Elith, J. & Wintle, B.A. (2017). Evaluating 318 continental-scale species distribution models over a 60-year prediction horizon: what factors influence the reliability of predictions? *Glob. Ecol. Biogeogr.*, 26, 371–384.
- Murray, J.V., Low Choy, S., McAlpine, C.A., Possingham, H.P. & Goldizen, A.W. (2011). Evaluating model transferability for a threatened species to adjacent areas: implications for rock-wallaby conservation. *Austral Ecol.*, 36, 76–89.
- Norberg, A., Abrego, N., Blanchet, F.G., Adler, F.R., Anderson, B.J., Anttila, J. *et al.* (2019). A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. *Ecol. Monogr.*, 89, e01370.
- Parravicini, V., Azzurro, E., Kulbicki, M. & Belmaker, J. (2015). Niche shift can impair the ability to predict invasion risk in the marine realm: an illustration using Mediterranean fish invaders. *Ecol. Lett.*, 18, 246–253.
- Pearman, P.B., Guisan, A., Broennimann, O., Randin, C.F., Peter, B., Guisan, A. *et al.* (2008). Niche dynamics in space and time. *Trends Ecol. Evol.*, 23, 149–158.
- 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. *Glob. Ecol. Biogeogr.*, 26, 275–287.
- Petitpierre, B., Kueffer, C., Broennimann, O., Randin, C., Daehler, C. & Guisan, A. (2012). Climatic niche shifts are rare among terrestrial plant invaders. *Science*, 335, 1344–1348.
- Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006). Maximum entropy modeling of species geographic distributions. *Ecol. Modell.*, 190, 231–259.
- Qiao, H., Feng, X., Escobar, L.E., Peterson, A.T., Soberón, J., Zhu, G. *et al.* (2019). An evaluation of transferability of ecological niche models. *Ecography*, 42, 521–534.
- Qiao, H., Soberón, J. & Peterson, A.T. (2015). No silver bullets in correlative ecological niche modelling: insights from testing among many potential algorithms for niche estimation. *Methods Ecol. Evol.*, 6, 1126–1136.
- R Development Core Team. (2018). *R: a language and environment for statistical computing: R Foundation for Statistical Computing*. Austria, Vienna. Available at: <http://Rproject.org>.
- Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillerá-Arroita, G. *et al.* (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40, 913–929.
- Saul, W.C., Roy, H.E., Booy, O., Carnevali, L., Chen, H.J., Genovesi, P. *et al.* (2017). Assessing patterns in introduction pathways of alien species by linking major invasion data bases. *J. Appl. Ecol.*, 54, 657–669.
- Seebens, H., Blackburn, T.M., Dyer, E.E., Genovesi, P., Hulme, P.E., Jeschke, J.M. *et al.* (2018). Global rise in emerging alien species results from increased accessibility of new source pools. *Proc. Natl Acad. Sci.*, 115, E2264–E2273.
- Sequeira, A.M.M., Bouchet, P.J., Yates, K.L., Mengersen, K. & Caley, M.J. (2018). Transferring biodiversity models for conservation: opportunities and challenges. *Methods Ecol. Evol.*, 9, 1250–1264.
- Soininen, J., Korhonen, J.J. & Luoto, M. (2013). Stochastic species distributions are driven by organism size. *Ecology*, 94, 660–670.
- Soininen, J. & Luoto, M. (2014). Predictability in species distributions: a global analysis across organisms and ecosystems. *Glob. Ecol. Biogeogr.*, 23, 1264–1274.
- Strubbe, D., Beauchard, O. & Matthysen, E. (2015). Niche conservatism among non-native vertebrates in Europe and North America. *Ecography*, 38, 321–329.
- Thuiller, W., Guéguen, M., Renaud, J., Karger, D.N. & Zimmermann, N.E. (2019). Uncertainty in ensembles of global biodiversity scenarios. *Nat. Commun.*, 10, 1446.
- Thuiller, W., Lavorel, S. & Araújo, M.B. (2005). Niche properties and geographical extent as predictors of species sensitivity to climate change. *Glob. Ecol. Biogeogr.*, 14, 347–357.
- Tingley, R., Thompson, M.B., Hartley, S. & Chapple, D.G. (2016). Patterns of niche filling and expansion across the invaded ranges of an Australian lizard. *Ecography*, 39, 270–280.
- Václavík, T. & Meentemeyer, R.K. (2012). Equilibrium or not? Modelling potential distribution of invasive species in different stages of invasion. *Divers. Distrib.*, 18, 73–83.
- VanDerWal, J., Shoo, L.P., Graham, C. & Williams, S.E. (2009). Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? *Ecol. Modell.*, 220, 589–594.
- Vilà, M. & Hulme, P.E. (2017). *Impact of biological invasions on ecosystem services*. Springer, New York, NY.

- Wiens, J.A., Stralberg, D., Jongsomjit, D., Howell, C.A. & Snyder, M.A. (2009). Niches, models, and climate change: assessing the assumptions and uncertainties. *Proc. Natl Acad. Sci.*, 106, 19729–19736.
- Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A. *et al.* (2008). Effects of sample size on the performance of species distribution models. *Divers. Distrib.*, 14, 763–773.
- Yackulic, C.B., Chandler, R., Zipkin, E.F., Royle, J.A., Nichols, J.D., Campbell Grant, E.H. *et al.* (2013). Presence-only modelling using MAXENT: when can we trust the inferences? *Methods Ecol. Evol.*, 4, 236–243.
- Yates, K.L., Bouchet, P.J., Caley, M.J., Mengersen, K., Randin, C.F., Parnell, S. *et al.* (2018). Outstanding challenges in the transferability of ecological models. *Trends Ecol. Evol.*, 33, 790–802.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Editor, Cleo Bertelsmeier

Manuscript received 17 March 2020

First decision made 11 May 2020

Second decision made 24 June 2020

Manuscript accepted 30 June 2020