

<https://helda.helsinki.fi>

Occurrence-habitat mismatching and niche truncation when modelling distributions affected by anthropogenic range contractions

Pang, Sean E. H.

2022-06

Pang , S E H , Zeng , Y , De Alban , J D T & Webb , E L 2022 , ' Occurrence-habitat mismatching and niche truncation when modelling distributions affected by anthropogenic range contractions ' , Diversity and Distributions , vol. 28 , no. 6 , pp. 1327-1343 . <https://doi.org/10.1111/ddi.13544>

<http://hdl.handle.net/10138/344552>

<https://doi.org/10.1111/ddi.13544>

cc_by

publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Occurrence–habitat mismatching and niche truncation when modelling distributions affected by anthropogenic range contractions

Sean E. H. Pang¹  | Yiwen Zeng^{1,2}  | Jose Don T. De Alban^{1,2}  | Edward L. Webb^{1,3,4} 

¹Department of Biological Sciences, National University of Singapore, Singapore, Singapore

²Centre for Nature-Based Climate Solutions, National University of Singapore, Singapore, Singapore

³Department of Forest Sciences, Viikki Tropical Resources Institute, University of Helsinki, Helsinki, Finland

⁴Helsinki Institute of Sustainability Science (HELSUS), Helsinki, Finland

Correspondence

Sean E. H. Pang, Department of Biological Sciences, National University of Singapore, Blk S3 Level 1, Science Drive 4, Singapore City 117543, Singapore.
Email: s.pang@u.nus.edu

Funding information

Ministry of Education - Singapore

Editor: Boris Leroy

Abstract

Aims: Human-induced pressures such as deforestation cause anthropogenic range contractions (ARCs). Such contractions present dynamic distributions that may engender data misrepresentations within species distribution models. The temporal bias of occurrence data—where occurrences represent distributions before (*past bias*) or after (*recent bias*) ARCs—underpins these data misrepresentations. Occurrence–habitat mismatching results when occurrences sampled before contractions are modelled with contemporary anthropogenic variables; niche truncation results when occurrences sampled after contractions are modelled without anthropogenic variables. Our understanding of their independent and interactive effects on model performance remains incomplete but is vital for developing good modelling protocols. Through a virtual ecologist approach, we demonstrate how these data misrepresentations manifest and investigate their effects on model performance.

Location: Virtual Southeast Asia.

Methods: Using 100 virtual species, we simulated ARCs with 100-year land-use data and generated temporally biased (*past* and *recent*) occurrence datasets. We modelled datasets with and without a contemporary land-use variable (conventional modelling protocols) and with a temporally dynamic land-use variable. We evaluated each model's ability to predict historical and contemporary distributions.

Results: Greater ARC resulted in greater occurrence–habitat mismatching for datasets with *past bias* and greater niche truncation for datasets with *recent bias*. Occurrence–habitat mismatching prevented models with the contemporary land-use variable from predicting anthropogenic-related absences, causing overpredictions of contemporary distributions. Although niche truncation caused underpredictions of historical distributions (environmentally suitable habitats), incorporating the contemporary land-use variable resolved these underpredictions, even when mismatching occurred. Models with the temporally dynamic land-use variable consistently outperformed models without.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Diversity and Distributions* published by John Wiley & Sons Ltd.

Main conclusions: We showed how these data misrepresentations can degrade model performance, undermining their use for empirical research and conservation science. Given the ubiquity of ARCs, these data misrepresentations are likely inherent to most datasets. Therefore, we present a three-step strategy for handling data misrepresentations: maximize the temporal range of anthropogenic predictors, exclude mismatched occurrences and test for residual data misrepresentations.

KEYWORDS

data quality, ecological niche model, habitat loss, land cover change, model performance, niche reduction, species distribution model, temporal misalignment, temporal sampling bias

1 | INTRODUCTION

As global changes increasingly threaten biodiversity and ecosystems, there is a need to model and understand the effects of these changes on species distributions (Guisan & Thuiller, 2005; Guisan et al., 2013; Newbold, 2018). The most extensive and prevalent changes in recent times are caused by human pressures such as poaching or deforestation, which manifests in anthropogenic range contractions (ARCs) (Maxwell et al., 2016; Newbold et al., 2015). Hence, predictions of biodiversity scenarios need to integrate species responses to such pressures through anthropogenic variables such as land-use/land-cover (Titeux et al., 2016).

A common approach towards measuring and predicting changes in species distributions are species distribution models (SDMs), which estimate habitat suitability by correlating species' occurrences to prevailing environmental conditions (Soberon & Peterson, 2005). However, SDMs typically assume distributions to be at equilibrium with their environment (Araújo & Pearson, 2005), which may lead to data misrepresentations when modelling species already affected by ARCs. The crux of these data misrepresentations stems from the incongruence between static occurrence data and temporally dynamic

distributions, where occurrence datasets will inevitably exhibit some form of temporal bias relative to the progression of ARC (Boakes et al., 2010; Ryo et al., 2019). Datasets biased to the *past* or to more *recent* times, respectively, are more likely to have occurrences sampled before or after an ARC.

Consider the occurrence records of a forest-dependent species sampled from forests in the decades prior to clearance in the year 1980 for urban land-use (Figure 1). When an SDM using those occurrence data incorporate a land-use variable from the year 2000, occurrence-habitat mismatching occurs, as occurrences are mismatched with the unsuitable urban land-use class (Milanesi et al., 2020; Pang et al., 2021; Ryo et al., 2019). Thus, the resultant model may wrongly infer urban land-use as suitable and underestimate the true impact of deforestation on the species' range. This situation arises because of the lack of temporal range and resolution for many anthropogenic variables, and conventional use of static predictors in most SDM studies; hence, contemporary anthropogenic variables are often utilized in SDM studies despite possible mismatching with historical occurrences (Garcia et al., 2013; Marshall et al., 2018; Seaborn et al., 2021). Although modelling without anthropogenic variables avoids potential mismatching, where range contractions

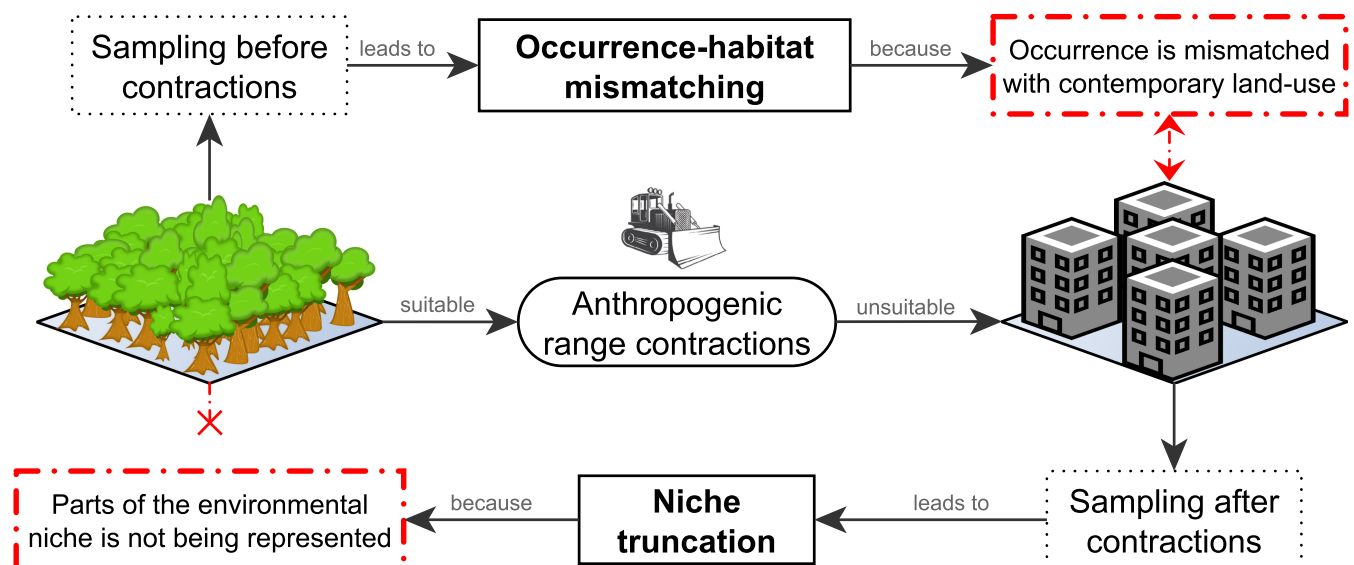


FIGURE 1 Concept chart illustrating the processes through which differential sampling across time, relative to when anthropogenic range contraction had occurred, can lead to occurrence-habitat mismatching or niche truncation within an occurrence dataset

can be accounted for post-hoc (see Gomes et al., 2019; Manhego et al., 2017; Newbold, 2018), excluding anthropogenic variables facilitates niche truncation when occurrences are sampled after an ARC.

Consider alternatively, the same species and deforestation scenario, but wherein occurrences were sampled in 1990, that is, after land-use conversions (Figure 1). Those occurrences would represent only a subset of the species' historical range, and by extension, a subset of the species' historically realized niche (Colwell & Rangel, 2009; Hutchinson, 1957; Scheele et al., 2017). When an SDM using those occurrence data lack a land-use variable, niche truncation occurs, because absences due to urban land-use are being erroneously attributed to the available environmental variables (Barve et al., 2011; Owens et al., 2013). This truncated estimate of the niche may undermine model transferability, leading to inaccurate predictions of species distribution across space (e.g. invasive potential) and time (e.g. climate change) (Guisan et al., 2014; Peterson et al., 2018). Occurrence-habitat mismatching and niche truncation are thus conceptually related outcomes of ARCs but arise out of opposing temporal biases and modelling protocols.

Despite extensive research into misrepresentations of suitable habitats due to a suite of factors (Chen et al., 2012; Guélat & Kéry, 2018; Guillera-Aroita, 2017; Soley-Guardia et al., 2016), occurrence-habitat mismatching due to temporal biases remains poorly explored, especially within the increasingly pervasive context of ARCs (Boakes et al., 2010; Milanese et al., 2020; Ryo et al., 2019). In comparison, multiple real species studies have found evidence for niche truncation due specifically to ARCs (Faurby & Araújo, 2018; Gibson et al., 2019; Martínez-Freiría et al., 2016; Rutrough et al., 2019). However, utilizing real species hinders assessments of model performance because knowledge on historical distributions and the full extent of ARCs are incomplete. Indeed, neither data misrepresentation can be accurately quantified for real species datasets, making it difficult to disentangle their effects on SDMs. Furthermore, because previous studies had investigated either data misrepresentation in isolation, possible interactions between them have gone unstudied. Specifically, although studies have shown that incorporating an anthropogenic variable allows SDMs to account for niche truncation (Gibson et al., 2019; Nüchel et al., 2018; Requena-Mullor et al., 2019; Silva et al., 2012), occurrence-habitat mismatching may undermine the effectiveness of such an approach. Developing effective solutions for overcoming these data misrepresentations requires a thorough understanding of their independent and interactive effects on SDM performance (Araújo et al., 2019); in other words, evaluating these data misrepresentations in tandem.

In this study, we explore how data misrepresentations may arise from interactions between ARCs and temporal sampling bias and evaluate them simultaneously to isolate their independent and interactive effects on SDM performance. We overcome limitations associated with incomplete knowledge on historical distributions and extent of ARC by adopting a virtual ecologist approach (Austin et al., 2006; Meynard et al., 2019; Zurell et al., 2010). Using 100 virtual species, we simulated ARCs with real anthropogenic land-use

data, and generated occurrence datasets with temporal biases based on real-world trends. First, we quantify and characterize these data misrepresentations as a function of both ARC and the temporal bias. Second, we model datasets with and without a contemporary anthropogenic variable and with a temporally dynamic anthropogenic variable, before comparing model predictions against the true historical and contemporary distributions of our virtual species. Based on the findings, we then propose strategies to tackle these data misrepresentations and improve model reliability.

2 | METHODS

The study encompasses Southeast Asia west of Wallace's Line for the period 1900–2000. The workflow was divided into six phases: (1) create virtual species; (2) simulate ARCs; (3) generate temporally explicit occurrence datasets; (4) model virtual species distributions; (5) evaluate model performance and (6) quantify data misrepresentations and their effects on model performance (Figure 2) (for detailed workflow, see Supporting Information Figure S1.1).

2.1 | Create virtual species

The virtual ecologist approach works by simulating species distributions and observer models to generate virtual occurrence data (Zurell et al., 2010). The virtual occurrence data can be modelled and evaluated against the 'true' distribution, thus supporting an assessment of factors contributing to data quality, and a quantification of SDM performance (Austin et al., 2006; Meynard et al., 2019). The true distribution of virtual species, henceforth referred to as species unless otherwise specified, was based on environmental suitability and fractional cover of anthropogenic land-use. Environmental suitability was determined by the species' environmental niche, which was created from 15 bioclimatic (5 arc min resolution) and nine soil variables (250 m resolution) sourced from WorldClim.org and SoilGrid.org respectively (Fick & Hijmans, 2017; Hengl et al., 2017) (Table S1.2). These variables were first resampled (bilinear) to 5 arc min resolution before a principal component analysis was conducted, where the first five PC-axes (~85% variance) were retained (Table S1.3). Retained PC-axes were then used to randomly create environmental niches through the 'generateRandomSp' function from the 'virtualespecies' package in R, which produced continuous estimates of environmental suitability in geographical space based on realistically viable responses (for details see Leroy et al., 2016). One hundred environmental niches and their corresponding estimates of environmental suitability were generated, which were allowed a maximum pairwise Pearson's correlation coefficient of .85 to prevent duplicate results. Although distributional shifts due to intra-/inter-annual climate variability may also result in misrepresentation of suitable habitats (Milanese et al., 2020), we kept environmental variables static across the entire study period as this study focused on distributional changes caused by anthropogenic pressures.

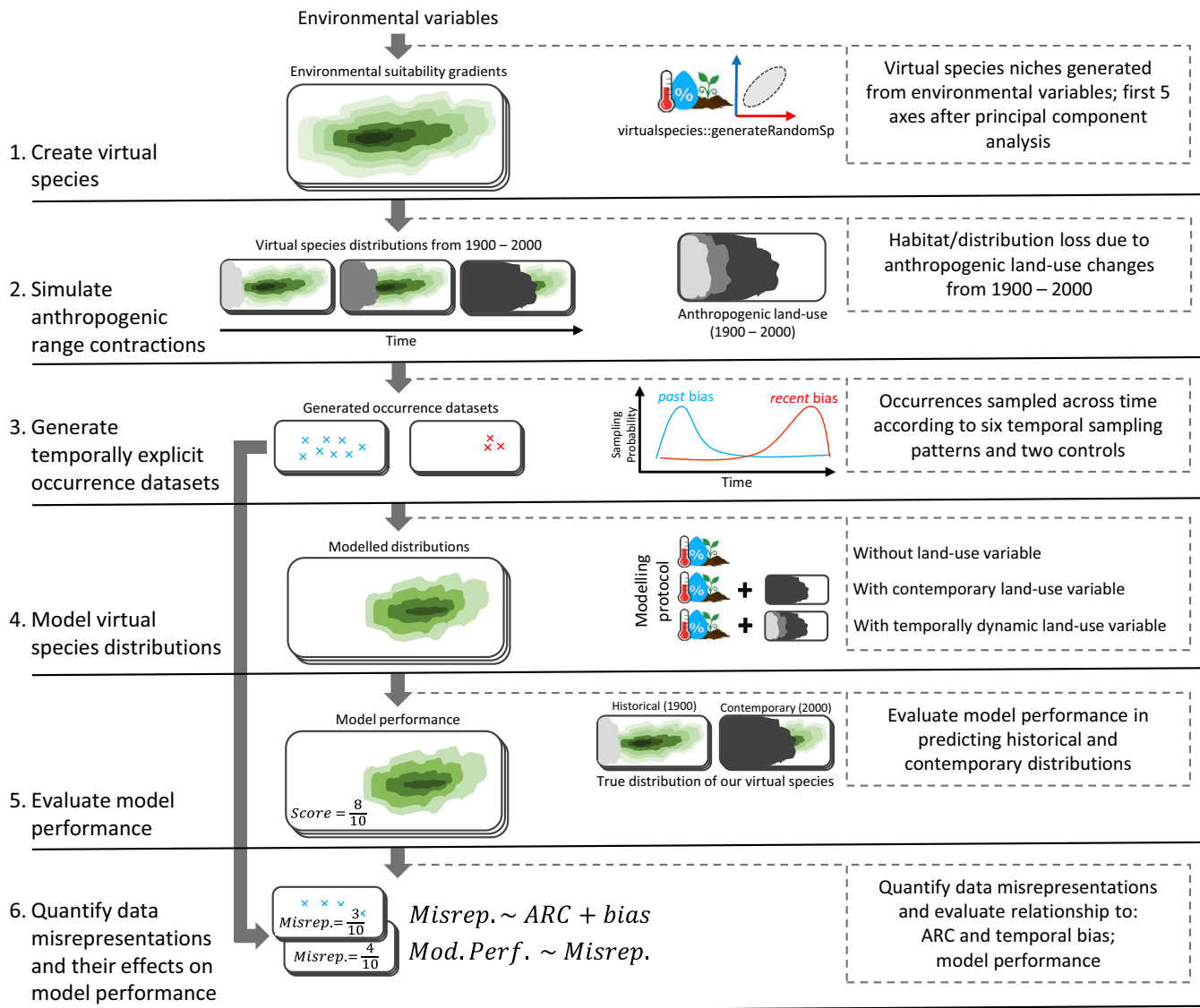


FIGURE 2 Methods flowchart for this study indicating the six phases of this study: 1) create virtual species; 2) simulate anthropogenic range contractions; 3) generate temporally explicit occurrence datasets; 4) model virtual species distributions; 5) evaluate model performance and 6) quantify data misrepresentations and their effects on model performance. A more detailed version can be found in Figure S2.1

2.2 | Simulate anthropogenic range contractions

Annual land-use maps (0.25° resolution rasters) from 1900 to 2000 were obtained from the Land Use Harmonization dataset (LUH2 v2h; <<https://luh.umd.edu/data.shtml>>), which indicated fractional cover for several classes (Hurt et al., 2020). We generated an anthropogenic land-use class by summing urban, rangelands, managed pastures, C3 and C4 annual crops, C3 and C4 perennial crops, and C3 nitrogen-fixing crops land-use classes for each year. The maps were resampled (bilinear) to 5 arc min resolution to match environmental variables. Fractional cover of anthropogenic land-use was the only land-use class considered for this study, which was simulated to negatively affect habitat suitability. The negative effect was defined by a negative logistic regression that interacted with environmental suitability on a multiplicative scale, which replicated the contagion-like

spread of extinction forces observed to induce ARCs (Figure S1.4) (Channell & Lomolino, 2000). This produced the final, annual, true distribution maps of continuous habitat suitability for the 100 species, based on their environmental niche and the fractional cover of anthropogenic land-use. The continuous true distribution maps of habitat suitability for each species were then converted to binary presence–absence distributions using species-specific thresholds of habitat suitability. Using the ‘convertToPA’ function from the ‘virtualspecies’ package in R (Leroy et al., 2016; R Core Team, 2013), thresholds were determined by fixing the initial prevalence (year 1900) of species at 0.25. Additionally, four other prevalence values were used to test for sensitivity (Figure S2.3–S2.6).

Anthropogenic range contraction (proportion of initial range lost) for a species was calculated as the proportion of ‘presence’ pixels that changed to ‘absence’ from 1900 to 2000, due to an increase

in fractional cover of anthropogenic land-use. We utilized the 'raster' package in R for the resampling of variables and the summation of the different land-use classes (Hijmans & Etten, 2012; R Core Team, 2013).

2.3 | Generate temporally explicit occurrence datasets

We simulated six temporal sampling patterns—two temporal sampling distributions and three temporal sampling biases—based on generalizations of real species occurrence datasets (Figure 3; Figures S1.5 and S1.6) (GBIF.org, 2020). This ensured realistically possible data misrepresentations within each generated occurrence dataset. Temporal sampling distributions were either *clustered* (sampled over a short period) or *spread* (sampled more evenly across time). Temporal sampling biases were either *past* (majority sampled near the start of the study period), *recent* (near the end), or *intermediate* (neither near the start nor end). Two extreme sampling patterns were also included to act as both positive and negative controls: *only start* (all occurrences sampled in 1900) served as the positive control for occurrence–habitat mismatching and the negative control for niche truncation; and vice versa for *only end* (all occurrences sampled in 2000). Positive controls should maximize their respective data misrepresentations.

To generate temporally explicit species occurrence datasets (i.e. presence-only), 2000 points were randomly sampled across the study region. For each point, the probability of sampling from a

particular year was based on the simulated sampling frequencies of a temporal sampling pattern (Figure 3). This was repeated for each temporal sampling pattern (including controls), which generated eight temporally unique datasets. Lastly, the occurrence probability for each sampling point was derived from the suitability of the habitat (both environment and land-use) and the binary range of the species for that year; occurrence probability outside the range was always zero. As replicates, this entire process was repeated 10 times. In total, 8000 occurrence datasets were generated (8 temporal sampling patterns × 100 species × 10 replicates).

2.4 | Model virtual species distributions

Species distributions models were developed for each occurrence dataset using the MaxEnt algorithm with its default settings (Phillips et al., 2006). The SDMs included the same environmental variables used to generate species as predictors (first five PC-axes), with 10,000 randomly sampled background points (Chefaoui & Lobo, 2008; VanDerWal et al., 2009). Models were trained under three protocols: without a land-use predictor, with a contemporary land-use predictor, and with a temporally dynamic land-use predictor. Land-use predictors here were the same fractional cover of anthropogenic land-use used to simulate ARC. Models using a contemporary land-use predictor had occurrence and background points associated with land-use values from the year 2000, whereas models using a temporally dynamic land-use predictor had occurrence and background points associated with land-use values from the

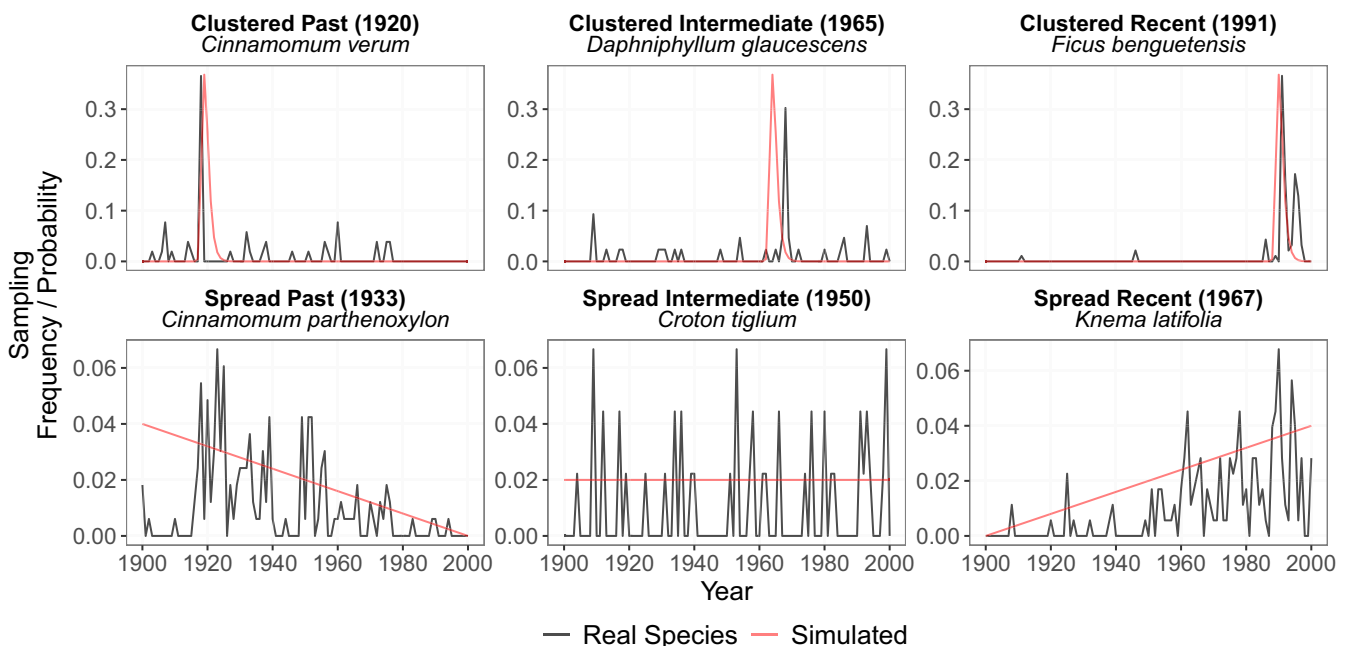


FIGURE 3 The six temporal sampling patterns (and median sampling year) visualized using sampling frequency (relative to the total number of occurrences for that species' dataset); simulated sampling frequencies in red and observed sampling frequencies for an example species in black. Patterns followed either a *clustered* or *spread* distribution (rows) and either a *past*, *intermediate*, or *recent* bias (columns). The simulated sampling frequencies for the six temporal sampling patterns were used to determine the probability of sampling from a specific year. Sampling frequency/probability for the controls (not shown) was 1 for their respective year

year of their collection (background points were temporally sampled to match the temporal sampling pattern of the occurrence dataset). For models trained with a land-use predictor, the land-use variable for the year 2000 was used to predict contemporary distributions and for the year 1900 to predict historical distributions. To delineate species' ranges, the probability distributions were converted to binary presence-absence using the maximizing sum of sensitivity and specificity threshold (Liu et al., 2005). All models were built with the 'sdmtune' package using the 'maxnet' function in R (Phillips et al., 2017; Vignali et al., 2019).

2.5 | Evaluate model performance

To evaluate model performance, resultant probability and binary distribution maps were compared against the true contemporary (year 2000) and historical (year 1900) distributions. Because models without a land-use predictor produced only one output, the same modelled distribution was compared against the true contemporary and historical distributions. For probability distributions, area under the curve (AUC) was used to evaluate discriminatory accuracy (Fielding & Bell, 1997), and Pearson correlation for functional accuracy (Warren et al., 2020) (Table 1). Functional accuracy refers to a model's ability to predict relative suitability values correctly, which relates to the reliability of biological inferences made based on those values (Warren et al., 2020). To identify the specific binary errors in model predictions, underprediction rates (i.e. omission rate or false negative rate) and overprediction rates (i.e. commission rate or false positive rate) were calculated (Leroy et al., 2018). While the underprediction rate measures the proportion of true presence falsely predicted by the model, the overprediction rate measures the proportion of predicted presence falsely predicted by the model (Table 1). For comparison, we also calculated two other common performance metrics, true skill statistic (TSS) and Cohen's kappa (Allouche et al., 2006) (Figure S2.4–S2.6).

2.6 | Quantify data misrepresentations and their effects on model performance

We used two indices to quantify occurrence-habitat mismatching and niche truncation within each occurrence dataset: mean difference in habitat suitability and niche dissimilarity respectively. For each occurrence point, a mismatch between the occurrence and associated habitat was calculated by subtracting the true suitability of the habitat for the year 2000 from its true suitability at the time of sampling, that is, the difference in habitat suitability. Difference values were then averaged across points to quantify occurrence-habitat mismatching within a dataset.

Niche dissimilarity was calculated as $1 - I$, where I is the overlap between niches represented by the true historical distribution (year 1900) and occurrence dataset for each species (Warren et al., 2008). The niches were derived from environmental variables only (five

PC-axes), using the 'nicheOverlap' function from the 'dismo' package in R (Hijmans et al., 2017; R Core Team, 2013). The niche dissimilarity score from the 10 negative controls for each species served as a baseline value for species-specific correction (Figure S1.7). Because of this correction, negative values resulted, which indicates that the calculated niche truncation was less than the established baseline.

The two indices, mean difference in habitat suitability and niche dissimilarity (hereafter simply occurrence-habitat mismatching and niche truncation), were plotted against ARC (proportion of range lost) for each temporal sampling pattern. Generalized linear models (GLMs) were used to assess the effects of data misrepresentations on the performance of SDMs under the three modelling protocols. GLMs were built separately to evaluate model predictions of contemporary and historical distributions, for each model evaluation metric. The GLMs were built in R using the 'stats' package (R Core Team, 2013).

3 | RESULTS

3.1 | The manifestation of occurrence-habitat mismatching and niche truncation

Our results showed that the proportion of range lost during ARC was positively correlated to both data misrepresentations (Figure 4). This was clearest for the positive controls, which maximized their respective misrepresentations and had regression lines against ARC with the highest slope and R^2 values (slope = 0.352, $R^2 = .73$ for occurrence-habitat mismatching; slope = 0.232, $R^2 = .44$ for niche truncation). Regression slopes, however, varied among temporal sampling biases: datasets with a *past* bias exhibited a higher slope for occurrence-habitat mismatching but a lower slope for niche truncation and vice versa for datasets with a *recent* bias (Figure 4). This demonstrates ARC as the primary driver of both data misrepresentations, while the temporal sampling bias determined the specific data misrepresentation that manifests. We also observed a significant negative partial correlation between data misrepresentations when controlling for ARC ($n = 8000$, $r = -.63$, $p < 0.001$), indicating a negative relationship between them. For clarity and simplicity, *clustered* and *spread* temporal sampling distributions were aggregated since trends between them were similar (for disaggregated results, see Figure S2.1).

3.2 | Relationship between model performance and data misrepresentations

The effects of occurrence-habitat mismatching and niche truncation on model performance depended on the prediction scenario—that is, historical or contemporary distributions—and the modelling protocol. For predictions of historical distributions (Figure 5a), models without the land-use predictor were negatively affected by niche truncation. As niche truncation increased, predictions of models

TABLE 1 A summary table of the different evaluation metrics considered. For the evaluation of binary predictions: True Positives (TP) is the number of correctly predicted presence, False Positives (FP) is the number of falsely predicted presence, False Negatives (FN) is the number of falsely predicted absence and True Negatives (TN) is the number of correctly predicted absence; where (TP + FN) equals the true presence, (TP + FP) equals the predicted presence, and (TP + FP + FN + TN) equals the total number of sites (n)

Evaluates	Metric	Acronym	Definition/Formula
Probabilistic Accuracy	Area under the curve	AUC	Plot sensitivity against specificity - 1 and calculate the area under the curve
	Functional accuracy Pearson	FAP	Correlation between the true and modelled probability distributions, measures the accuracy of relative habitat suitability
Binary Error Rates	Underprediction rates	UR	Proportion of true presence falsely predicted by the model; $\frac{FN}{TP + FN}$
	Overprediction rates	OR	Proportion of predicted presence falsely predicted by the model; $\frac{FP}{TP + FP}$
Binary Accuracy	True Skill Statistic	TSS	Sum of model sensitivity and specificity; $\frac{TP}{TP + FN} + \frac{TN}{FP + TN} - 1$
	Cohen's kappa	KAPPA	Overall accuracy corrected against the expected accuracy; $\frac{\left(\frac{TP + TN}{n}\right) - \frac{(TP + FP)(TP + FN) + (FN + TN)(FP + TN)}{n^2}}{\frac{(TP + FP)(TP + FN) + (FN + TN)(FP + TN)}{n^2}}$

without the land-use predictor were less correlated with historical distributions and more severely underpredicted them. Models with either land-use predictor, however, were relatively unaffected by increases in either data misrepresentation; model correlation scores remained high while under- and overprediction rates remained low. Notably, for models with the contemporary land-use predictor, this meant niche truncation was accounted for despite occurrence-habitat mismatching and its effects. The spatially explicit example also confirmed these results. When niche truncation was high, underpredictions of the historical distribution only occurred for the model without the land-use predictor (Figure 6a). Although statistically significant, interaction between data misrepresentations were minor (for GLM coefficient estimates, see Figure S2.2).

On the other hand, for predictions of contemporary distributions (Figure 5b), reductions in model performance were observed primarily for models with and without the contemporary land-use predictor. For models without the land-use predictor, both data misrepresentations had a strong negative effect on model performance, whose increases resulted in predictions that were less correlated with contemporary distributions and more severely overpredicted them. This was because those models lacked the necessary predictor required to predict ARCs. For models with the contemporary land-use predictor, occurrence-habitat mismatching had strong negative and independent effects on model performance (lower correlation and higher overprediction). Niche truncation had no independent effects but interacted with occurrence-habitat mismatching to negatively affect model performance. However, this negative effect, like models without the land-use predictor, was due to model inability to predict ARCs. Specifically, as occurrence-habitat mismatching increased, model performance became increasingly similar to those observed for models without the land-use predictor, suggesting a convergence in model predictions. The spatially explicit example confirmed this convergence between models with and without the

contemporary land-use predictor, showing near-identical range estimates of contemporary distribution when occurrence-habitat mismatching was high (Figure 6b). Additionally, in response to a linear increase in occurrence-habitat mismatching, an exponential reduction in the variable importance of the contemporary land-use predictor was observed (Figure S2.7). While the exponential reduction in variable importance suggests model sensitivity to mismatching, the near zero variable importance implies an uninformative contemporary anthropogenic predictor.

Although occurrence-habitat mismatching was also found to negatively affect models with the temporally dynamic land-use predictor, the effects were substantially reduced under even minor increases in niche truncation (Figure 5b). This indicates that models with the temporally dynamic land-use variable generally performed well, except for predictions of contemporary distributions based on datasets almost entirely dominated by occurrence-habitat mismatching. This is because such datasets would not have captured any of the effects of ARC on species occurrences. Nevertheless, models with the temporally dynamic anthropogenic predictor typically outperformed the other two models, regardless of the prediction scenario or data misrepresentations present (Figures 5b and 6b).

3.3 | The insensitive AUC metric

Despite comparatively larger decreases in correlation scores and increases in either under- or overpredictions, the AUC metric displayed only minor reductions in model performance when either data misrepresentation increased (Figure 5a,b). This occurred across predictions of historical and contemporary distributions, and modelling protocols of land-use variables. Furthermore, this trend persisted for binary ranges based on other prevalence values (AUC > 0.80; Figure S2.4-S2.6) and model fit (i.e. AUC calculated

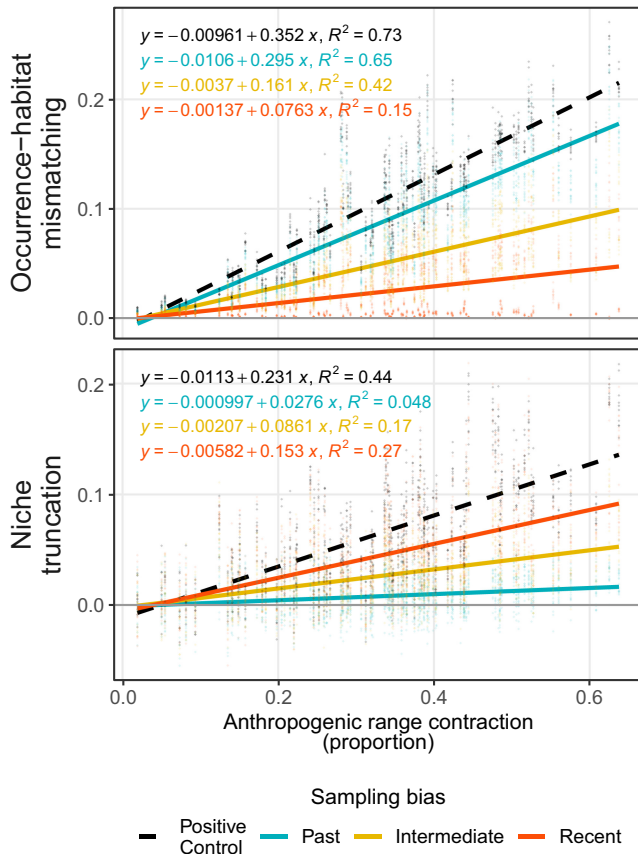


FIGURE 4 Linear regressions between anthropogenic range contractions and the two data misrepresentations, occurrence-habitat mismatching and niche truncation. Regression lines were disaggregated for each temporal sampling bias: past (blue), intermediate (yellow) and recent (red). Positive controls represent the theoretical maximum of each data misrepresentation (black dash): all occurrences sampled from the year 1900 (only start) for occurrence-habitat mismatching; and all occurrences sampled from the year 2000 (only end) for niche truncation. Anthropogenic range contraction was calculated as the proportion of the initial range (year 1900) lost by the end of the study period (year 2000). Note, niche truncation values here were after species-specific corrections, where a negative value indicates that the truncation was less than the established baseline

using the training data) ($AUC > 0.74$; Figure S2.8). This indicates that the AUC metric was relatively insensitive to model degradations due to occurrence-habitat mismatching and niche truncation. We did not find TSS nor KAPPA to be as insensitive to model degradations; TSS was insensitive to model overpredictions only (i.e. contemporary predictions), whereas reductions in KAPPA reflected decreases in correlation scores and increases in either under- or overpredictions (Figure S2.4–S2.6).

4 | DISCUSSION

Understanding the processes leading up to data misrepresentations and their subsequent effects on SDM performance is essential for developing good modelling practices (Fourcade et al.,

2018; Guillera-Arroita, 2017; Peterson et al., 2018; Yates et al., 2018). Yet, our understanding of occurrence-habitat mismatching and niche truncation, which arise out of interactions between ARCs and temporally biased data, remains incomplete. Importantly, their interactive effect on model performance is unknown. A complete understanding of these data misrepresentations is especially pertinent considering the extent and prevalence of ARCs and the unavoidability of temporally biased data (Boakes et al., 2010; Maxwell et al., 2016; Newbold et al., 2015). Through the virtual ecologist approach, we demonstrated how ARCs and temporal sampling biases can result in occurrence-habitat mismatching and niche truncation, and significantly impact model performance.

4.1 | Independent and (lack of) interactive effects on model performance

Using GLMs to evaluate the independent and interactive effect of data misrepresentations on model performance, two major findings emerged. First, occurrence-habitat mismatching impeded the modelling of species responses to the anthropogenic variable; that is, predicting anthropogenic-related absences. Owing to mismatching, models that included a contemporary anthropogenic variable overpredicted contemporary distributions in human modified landscapes, in a manner that was nearly identical to models without the anthropogenic variable. The anthropogenic variable was thus uninformative under circumstances of occurrence-habitat mismatching, specifically for predicting anthropogenic-related absences. In previous research, misrepresentations of environmentally suitable habitats due to vagrant individuals, imperfect detection, spatial biases or temporal biases, resulted in erroneous point-predictor associations (Chen et al., 2012; Guélat & Kéry, 2018; Guillera-Arroita, 2017; Milanese et al., 2020; Soley-Guardia et al., 2016). Although erroneous associations can skew variable importance, they seldom consistently and entirely reduce the informative capacity of the variable with which mismatching occurs (Harisena et al., 2021; Smith & Santos, 2020). With the present study, however, mismatching related specifically to the anthropogenic variable; as a result, no erroneous associations with environmental variables arose since mismatched occurrences remained accurate representations of environmental suitability. Rather, occurrence-habitat mismatching only confounded anthropogenic-related absences, thereby reducing the informative capacity of the anthropogenic variable that served strictly to predict those absences (Channell & Lomolino, 2000; Requena-Mullor et al., 2019). Second, in agreement with past studies, we found incorporating a contemporary anthropogenic variable to account for niche truncation (Gibson et al., 2019; Nüchel et al., 2018; Requena-Mullor et al., 2019; Silva et al., 2012); in addition, however, this also held true for models affected by occurrence-habitat mismatching.

Taken together, these findings suggest the use of the anthropogenic variable to predict anthropogenic-related absences and

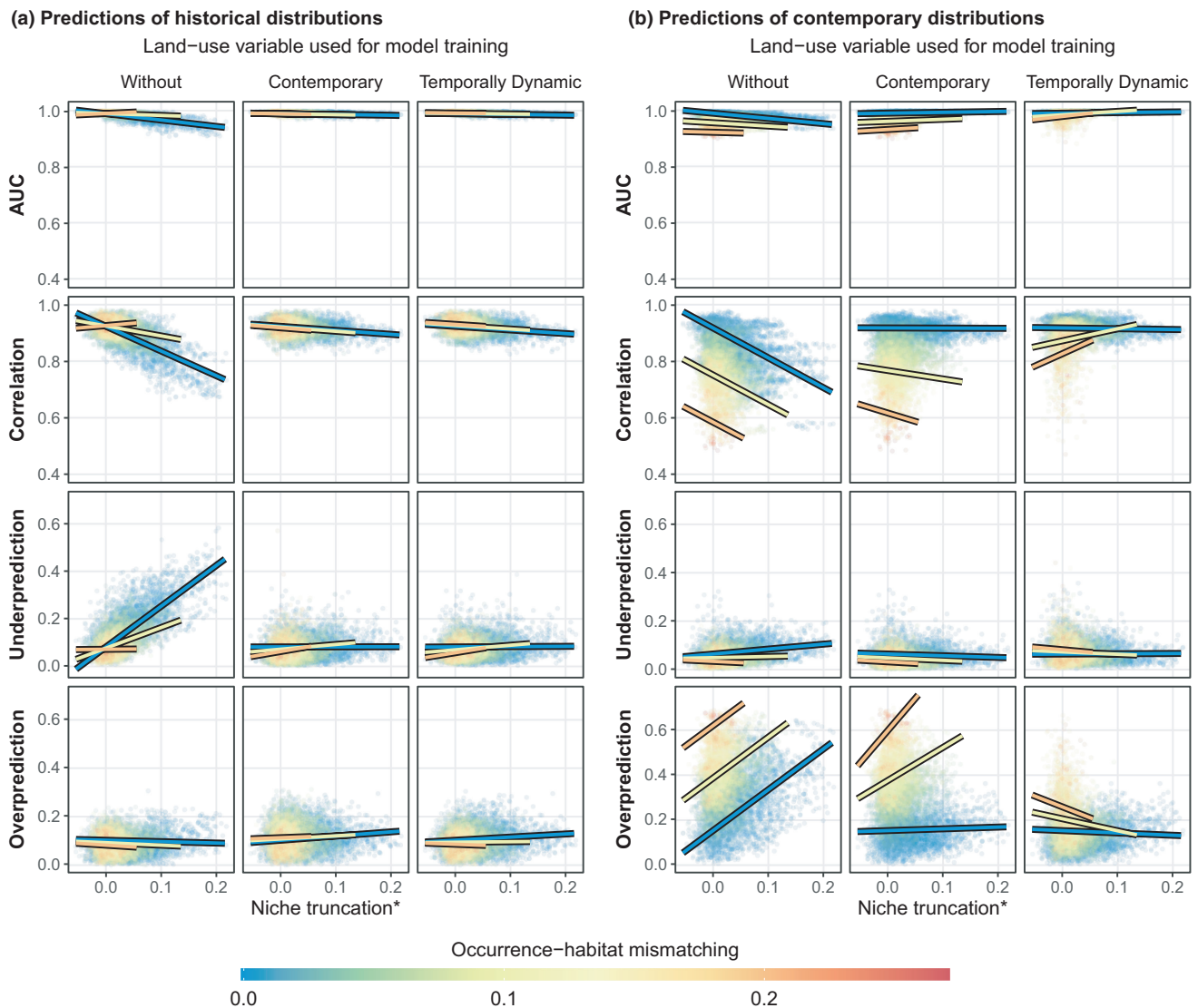


FIGURE 5 The plotted and modelled effects of occurrence–habitat mismatching and niche truncation on the performance of SDMs under different modelling protocols, for predictions of (a) historical and (b) contemporary distributions. Points represent the observed performance of SDMs, and regression lines indicate the estimated performance. Model performance was indicated by the y-axis, niche truncation by the x-axis, and occurrence–habitat mismatching by the spectral colours—specifically, 0.0 (blue), 0.1 (yellowish green) and 0.2 (orange) for the regression lines. Plots were faceted for each modelling protocol (horizontal facet) and metric of model performance (vertical facet). The metrics of model performance were for probabilistic performance (AUC and Pearson Correlation; higher is better) and binary error rates (Underprediction rates and Overprediction rates; lower is better). Note, niche truncation values here were after species-specific corrections, where a negative value indicates that the truncation was less than the established baseline

to account for niche truncation as independent processes; mismatching hinders the former but not the latter. This independence probably stems from the fact that range contractions due to anthropogenic pressures are spatially heterogeneous and can spread independently of environmental suitability and demographic processes (Channell & Lomolino, 2000; Scheele et al., 2017). As such, within the SDM itself, a species' response to the anthropogenic variable is modelled independently of the species' niche (Colwell & Rangel, 2009; Hutchinson, 1957; Scheele et al., 2017). Likewise, when modelling the niche, accounting for niche truncation through the anthropogenic variable is not based on the modelled response to that same variable.

4.2 | Implications for modelling ARC-affected species

The effects of occurrence–habitat mismatching have important methodological implications for modelling distributions shaped by anthropogenic pressures. In terms of modelling protocols, simply incorporating a contemporary anthropogenic variable into an SDM does not assure accurate predictions of distributions after ARC, regardless of the accuracy or precision of the anthropogenic variable, or its representativeness of the factors driving ARC. Should occurrence–habitat mismatching occur, models that incorporate a contemporary anthropogenic variable would severely underestimate

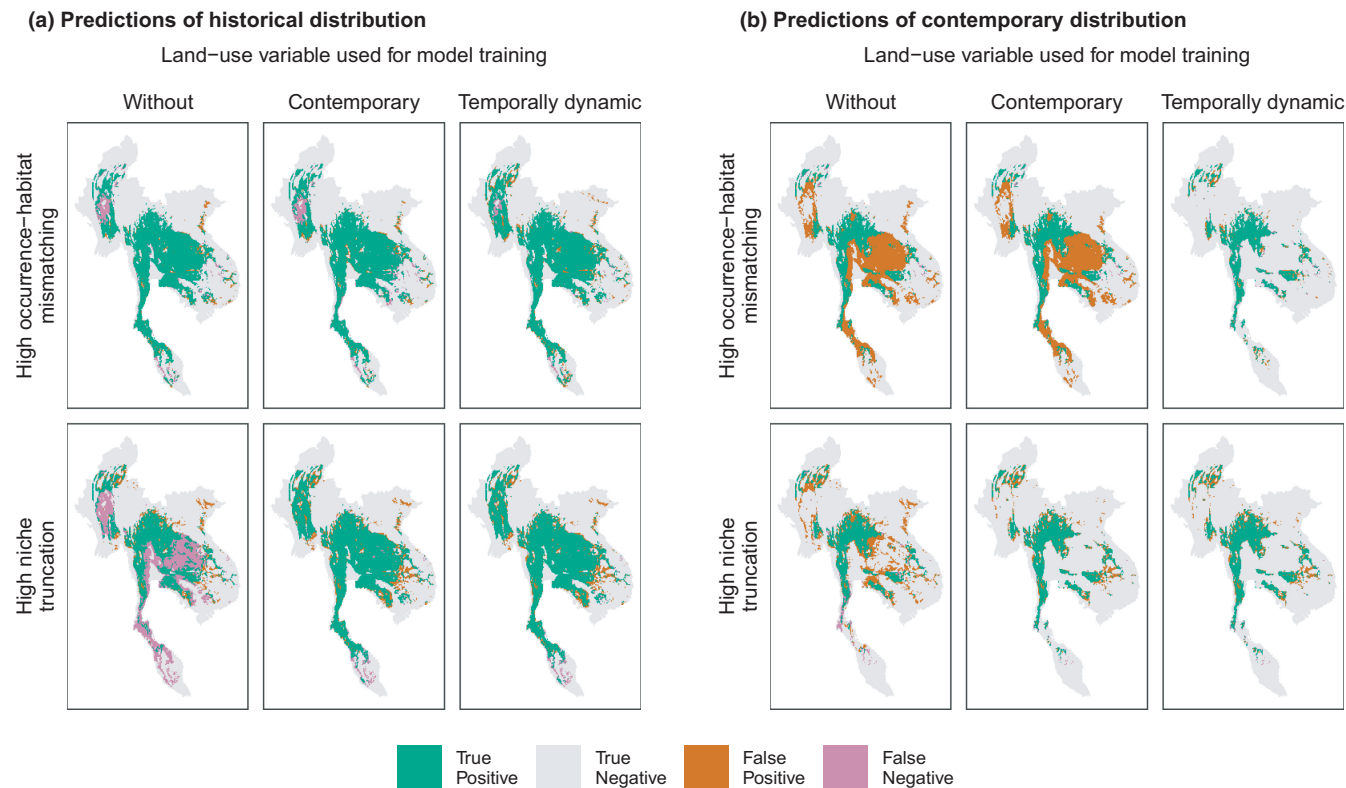


FIGURE 6 The spatially explicit binary accuracy of SDMs for predictions of an example species (a) historical and (b) contemporary distribution. Binary accuracy maps were faceted for each modelling protocol (horizontal facet) and for two different datasets (vertical facet). While one dataset exhibited high occurrence–habitat mismatching (0.25) but low niche truncation (0.02) (i.e. sampled following a *clustered past* temporal sampling pattern), the other exhibited high niche truncation (0.20) but low occurrence–habitat mismatching (0.00) (i.e. sampled following a *clustered recent* temporal sampling pattern). In this example, the selected species experienced a 62.5% contraction of its historical range (virtual species N27, see Supporting Information). For illustration purpose, only mainland Southeast Asia was shown as the main distribution of the selected species occurred there

the impact of human activity on species distributions. This could lead to overly optimistic range estimates that obscure the true threat statuses of species (Pang et al., 2021; Requena-Mullor et al., 2019; Titeux et al., 2016). The biased assessment of species vulnerabilities may then divert conservation efforts and funds away from species that need it (Fuller et al., 2010). Using resultant model outputs to inform systematic conservation planning can also skew habitat prioritizations and undermine the effective demarcation or expansion of protected areas, or worse, result in the allocation of resources into protecting already degraded habitats (Guisan et al., 2013; McShea, 2014; Struebig et al., 2015). Our result also suggests that models were sensitive to occurrence–habitat mismatching, where relatively minor mismatching could engender the problems described.

A key step towards ensuring model reliability is the detection of models affected by data misrepresentations (Araújo et al., 2019). Models affected by occurrence–habitat mismatching, however, will be challenging to detect. Without quality absence data, overpredictions are notoriously difficult to verify (Leroy et al., 2018; Warren et al., 2019). This detection difficulty is exacerbated by the uninformative outcome of the anthropogenic variable, in that multiple factors other than mismatching may lower variable importance as

well; whereas an erroneous response may be intuitively recognized (Gibson et al., 2019; Guevara et al., 2018; Soley-Guardia et al., 2016). It is also possible to misinterpret the uninformative anthropogenic variable as a legitimate response, resulting in false inference of species resilience. On a related note, the supposed uninformative contemporary anthropogenic variable is somewhat misleading since it could remain important for extrapolating truncated niches. This highlights a need to reconsider how we understand and evaluate variable importance within dynamic landscapes, especially for variables with atypical responses (i.e. constrain-only) (Fourcade et al., 2018; Harisena et al., 2021; Smith & Santos, 2020).

Our findings reveal the viability and robustness of incorporating contemporary anthropogenic variables as a solution to niche truncation arising from ARC. The importance of this is twofold. First, data and technological constraints often restrict anthropogenic variables to relatively contemporary timescales, datasets with large temporal ranges generally lack spatial resolution (e.g. the LUH dataset used in this study; Hurtt et al., 2020) or coverage (e.g. SERVIR-Mekong Land Cover; Saah et al., 2020). Thus, contemporary anthropogenic variables are often all that is available. Second, although the alternative—incorporating historical data to

reduce niche truncation—was done in past studies, those same studies recognized the limitation of such an approach; a complete niche cannot be guaranteed unless data from times predating early human civilization is obtained (Faurby & Araújo, 2018; Martínez-Freiría et al., 2016; Rutrough et al., 2019). Additionally, given the paucity and imprecisions of available occurrence records (Chen et al., 2012; Feeley & Silman, 2011; Ficetola et al., 2015; Goodwin et al., 2015; Meyer et al., 2016), obtaining sufficient decades-old yet accurate data to substantially reduce niche truncation is an improbable luxury for most species. Incorporating contemporary anthropogenic variables is therefore the most feasible approach to overcome niche truncation.

Our observed underpredictions of historical distributions due to niche truncations emphasize the wide-ranging implications of leaving truncations unaccounted. When predicting potentially suitable habitats based on truncated niches: reforestation or rewilding projects will struggle to identify suitable habitats lost due to anthropogenic pressures (Guisan et al., 2013; Jarvie & Svenning, 2018); invasive species models may fail to map potential dispersal corridors or areas of invasion (Barbet-Massin et al., 2018; Liu et al., 2020); and model forecasts under climate change would exaggerate rates of range shift or loss (Faurby & Araújo, 2018; Martínez-Freiría et al., 2016). We also found niche truncation to degrade the accuracy of continuous estimates of relative suitability. This implies erroneous response curves and skewed estimates of variable importance, making SDMs affected by niche truncation inappropriate for empirical applications in general (Guevara et al., 2018; Harisena et al., 2021; Smith & Santos, 2020; Warren et al., 2019). Furthermore, although several studies for or against niche conservatism have accounted for geographical niche truncations, few have accounted for truncations due to anthropogenic factors (Atwater et al., 2018; Guisan et al., 2014; Petitpierre et al., 2012; Zhu et al., 2017). In other words, previously observed deviations from niche conservatism may be an artefact of unresolved anthropogenic niche truncation instead. These examples show that anthropogenic niche truncation can lead to artefacts in SDM predictions, even in applications not focused on anthropogenic impacts.

4.3 | A recommended strategy for handling data misrepresentations

Based on our findings, we propose a general strategy for modeling species affected by ARC. The three-step strategy of Maximize, Exclude and Test (MET) (Figure 7).

Maximize the temporal range of anthropogenic variables without over-sacrificing spatial resolution and relevance to ARCs, among others. Practitioners will have options for representing the anthropogenic pressures that drive range contractions. Although our results support selecting anthropogenic variables that can maximize the temporal range of temporally dynamic models, we cannot neglect other variable qualities. High spatial resolution

may be prioritized over temporal range for regional/local SDMs, or species responses better explained by fine-scale variables (Mangiaccotti et al., 2013; Marshall et al., 2021). Alternatively, for ARCs due to other factors such as poaching, human-density estimates—regardless of its temporal range—may be more appropriate as a predictor than land-use/land-cover (Gibson et al., 2019). Therefore, it is important to select anthropogenic variables with large temporal ranges, while maintaining spatial resolutions aligned with the study objective(s) and representations of anthropogenic pressures relevant to the target species. Fortunately, the trade-off between temporal range and spatial resolution is being gradually reduced as the quality of anthropogenic variables like land-use/land-cover continue to improve (e.g. see Cao et al., 2021; Chen et al., 2020; Winkler et al., 2021).

Exclude mismatched occurrences when incorporating the selected anthropogenic variable to account for niche truncation. We demonstrated a negative relationship between occurrence-habitat mismatching and niche truncation; meaning, although problematic occurrence data are often removed (Guélat & Kéry, 2018; Guillera-Aroita, 2017; Soley-Guardia et al., 2016), excluding mismatched occurrences will exacerbate niche truncation. However, we also found that when models incorporated an appropriate anthropogenic variable, niche truncations were rendered relatively inconsequential. Hence, the recommended step of reducing occurrence-habitat mismatching while relying on the selected anthropogenic variable to account for niche truncation. Studies can identify these mismatches via inferred/informed species-habitat relationships or their time of sampling (e.g. a forest-dependent bird occurrence matched against an urban land-use class or sampled before the earliest anthropogenic variable and/or any major ARC). However, in cases where mismatched occurrence cannot be removed (e.g. too few occurrences or difficult to verify), species responses to anthropogenic pressures can still be accounted for post-hoc (see Gomes et al., 2019; Manchoغو et al., 2017; Newbold, 2018).

Test for residual data misrepresentations by validating model hindcasts and measuring variability in anthropogenic variable importance. Because of real-world data limitations, studies must verify the anthropogenic variable selected in step 'Maximize', and the approach used to exclude mismatched occurrences in step 'Exclude'. Model hindcasts—or projections onto a manually calibrated null/zero anthropogenic pressure predictor—can be validated using those previously excluded mismatched occurrences (Dobrowski et al., 2011; Maiorano et al., 2013), where high omission rates (underprediction rates) would indicate residual niche truncation unresolved by the anthropogenic variable (Boyce's index as a continuous alternative; Boyce et al., 2002; Hirzel et al., 2006). For species without mismatched occurrences, spatial and temporal cross-validations are a viable alternative (Roberts et al., 2017). Next, although difficult to detect models affected by occurrence-habitat mismatching, it is possible to flag likely candidates using cross-validation or resampling techniques (Nisbet et al., 2018; Roberts et al., 2017; Ryo et al., 2019). Model sensitivity to mismatching asserts that minor differences in

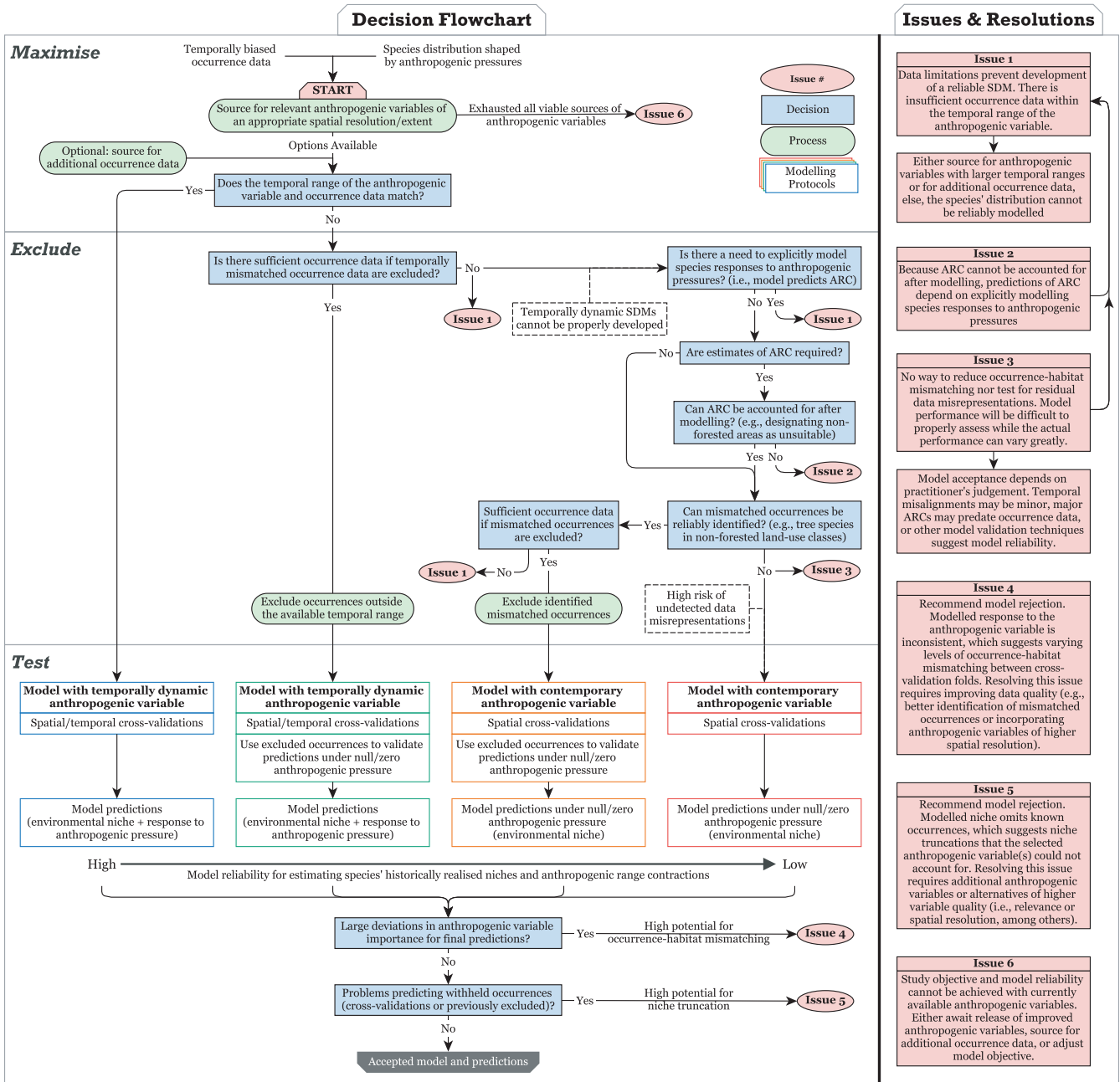


FIGURE 7 Decision flowchart detailing the three-step strategy of Maximize, Exclude and Test (MET), with four possible modelling protocols (differentiated by their outlined colour). Blue has the highest reliability since the temporal range of the anthropogenic variable matches that of the occurrence data. Green has high reliability as a temporally dynamic SDM; although the loss of data may affect model estimates of environmental niches, the validation of model hindcasts tests for niche truncation. Orange depends on the process used to exclude potential mismatched occurrences, which may be more reliable for range contractions driven by spatially observable pressures (e.g. deforestation as opposed to poaching). Red has the lowest reliability, which essentially bets on the incorporated anthropogenic variable to account for anthropogenic niche truncation

mismatching between partitions/resamples could result in highly varied anthropogenic variable importance. Large deviations in, or multiple outliers of, anthropogenic variable importance would therefore suggest residual occurrence-habitat mismatching. Ultimately, evidence of residual data misrepresentations should prompt reevaluations of steps 'Maximize' and 'Exclude', rejection of the model, or at least reporting and caveating model reliability.

4.4 | Study limitations

Our study was limited to the presence-background machine learning algorithm MaxEnt (Phillips et al., 2006, 2017). Hence, data misrepresentations might affect other statistical models differently, particularly those using dissimilar input data, such as demographic or presence-absence distribution models (Holden et al., 2021;

Norberg et al., 2019). However, statistical models using background or pseudo-absence points like MaxEnt can be expected to behave somewhat similarly but be more, or less, robust against data misrepresentations. Nevertheless, our limited testing of statistical models does not diminish the importance of recognizing occurrence–habitat mismatching and niche truncation and their implications for model performance.

Our findings may not apply to anthropogenic variables that integrate natural factors. For example, in representing deforestation, variables like canopy structure or vegetation indices. Since such variables may contain information relating to environmental suitability, mismatching with those variables may result in outcomes that differ from those observed in our study (Burns et al., 2020; Gibson et al., 2019; Girma et al., 2016). Moreover, while the anthropogenic land-use variable in this study perfectly represented our simulated ARCs, real-world anthropogenic variables will almost certainly have lower accuracy and precision (Prestele et al., 2016; Verburg et al., 2011), or exist at spatial scales too coarse to properly reflect species responses (Mangiacotti et al., 2013; Mertes & Jetz, 2018). Because of these real-world data constraints, incorporating a contemporary or temporally dynamic anthropogenic variable, in practice, may not result in model improvements comparable to those observed in our study.

5 | FUTURE PERSPECTIVES AND CONCLUSION

Our study reveals substantial challenges in modelling species distributions shaped by anthropogenic pressures. We showed how ARCs and temporally biased data can result in occurrence–habitat mismatching that impedes the predictions of anthropogenic-related absences, and niche truncations that lead to underestimations of environmentally suitable habitats. The effects of both data misrepresentations have important implications for SDM applications in various fields (Guisan et al., 2013; McShea, 2014; Peterson et al., 2018; Yates et al., 2018). These data misrepresentations may be underappreciated in part because of the relative insensitivity of the popular metric, AUC, which could have obscured the impacts of these data misrepresentations in past studies. Our findings also reveal that variables can be important on different fronts, which current methods seldom consider when evaluating variable importance (Elith et al., 2011; Harisena et al., 2021; Smith & Santos, 2020). Besides cross-validations and resampling techniques to examine deviations in variable importance, studies should consider measuring changes in predictive outcomes outside the training data, as well as within.

More research is equally required for developing and testing modelling protocols to overcome the challenges associated with modelling ARC-affected species. Our study highlighted the limitations of conventional modelling protocols; simply incorporating a contemporary anthropogenic variable or relying only on

environmental variables. Although our study demonstrated the potential of temporally dynamic SDMs (Milanesi et al., 2020), selecting appropriate variables is but one facet of developing good modelling practices (Araújo et al., 2019). Methods regarding the handling of presences and absences will be a logical next step, such as restricting background points to enable niche extrapolations without the anthropogenic variable (as hinted by in geographical niche truncation studies; see Barve et al., 2011; Owens et al., 2013; Saupe et al., 2012). For that matter, our recommended strategy aims to guide rather than dictate exact modelling processes and could serve as a precedent for future innovations. Crucially, with how ubiquitous ARCs are (Maxwell et al., 2016; Newbold et al., 2015), these data misrepresentations are likely the rule rather than the exception. Therefore, future studies should recognize occurrence–habitat mismatching and niche truncation borne from ARCs as highly pervasive and relevant and consider our recommended strategy for resolving them as an integral process of modelling species distributions.

ACKNOWLEDGEMENTS

We thank Ryan A. Chisholm for his comments, which greatly improved the final version of this manuscript. Funding for part of this research was from a Ministry of Education–Singapore AcRF Tier 1 Grant to ELW.

CONFLICT OF INTEREST

The authors of this study declare that they have no conflict of interest.

PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1111/ddi.13544>.

DATA AVAILABILITY STATEMENT

For reproducibility, all supporting data are archived in Dryad <<https://doi.org/10.5061/dryad.ttdz08m0g>> and the R scripts uploaded onto GitHub <<https://github.com/Pang-SEH/AnthropogenicRangeContractions-TemporalSamplingBias-SDM-Performance>>.

ORCID

Sean E. H. Pang  <https://orcid.org/0000-0003-4574-6028>

Yiwen Zeng  <https://orcid.org/0000-0001-7787-314X>

Jose Don T. De Alban  <https://orcid.org/0000-0002-1671-5786>

Edward L. Webb  <https://orcid.org/0000-0001-5554-9955>

REFERENCES

- Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS): Assessing the accuracy of distribution models. *Journal of Applied Ecology*, 43(6), 1223–1232. <https://doi.org/10.1111/j.1365-2664.2006.01214.x>
- Araújo, M. B., Anderson, R. P., Márcia Barbosa, A., Beale, C. M., Dormann, C. F., Early, R., Garcia, R. A., Guisan, A., Maiorano, L., Naimi, B., O'Hara, R. B., Zimmermann, N. E., & Rahbek, C. (2019). Standards

- for distribution models in biodiversity assessments. *Science Advances*, 5(1), eaat4858. <https://doi.org/10.1126/sciadv.aat4858>
- Araújo, M. B., & Pearson, R. G. (2005). Equilibrium of species' distributions with climate. *Ecography*, 28(5), 693–695. <https://doi.org/10.1111/j.2005.0906-7590.04253.x>
- Atwater, D. Z., Ervine, C., & Barney, J. N. (2018). Climatic niche shifts are common in introduced plants. *Nature Ecology & Evolution*, 2(1), 34–43. <https://doi.org/10.1038/s41559-017-0396-z>
- Austin, M. P., Belbin, L., Meyers, J. A., Doherty, M. D., & Luoto, M. (2006). Evaluation of statistical models used for predicting plant species distributions: Role of artificial data and theory. *Ecological Modelling*, 199(2), 197–216. <https://doi.org/10.1016/j.ecolmodel.2006.05.023>
- Barbet-Massin, M., Rome, Q., Villemant, C., & Courchamp, F. (2018). Can species distribution models really predict the expansion of invasive species? *PLoS One*, 13(3), e0193085. <https://doi.org/10.1371/journal.pone.0193085>
- Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S. P., Peterson, A. T., Soberón, J., & Villalobos, F. (2011). The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling*, 222(11), 1810–1819. <https://doi.org/10.1016/j.ecolmodel.2011.02.011>
- Boakes, E. H., McGowan, P. J. K., Fuller, R. A., Chang-qing, D., Clark, N. E., O'Connor, K., & Mace, G. M. (2010). Distorted views of biodiversity: Spatial and temporal bias in species occurrence data. *PLoS Biology*, 8(6), e1000385. <https://doi.org/10.1371/journal.pbio.1000385>
- Boyce, M. S., Vernier, P. R., Nielsen, S. E., & Schmiegelow, F. K. A. (2002). Evaluating resource selection functions. *Ecological Modelling*, 157(2), 281–300. [https://doi.org/10.1016/S0304-3800\(02\)00200-4](https://doi.org/10.1016/S0304-3800(02)00200-4)
- Burns, P., Clark, M., Salas, L., Hancock, S., Leland, D., Jantz, P., Dubayah, R., & Goetz, S. J. (2020). Incorporating canopy structure from simulated GEDI lidar into bird species distribution models. *Environmental Research Letters*, 15(9), 095002. <https://doi.org/10.1088/1748-9326/ab80ee>
- Cao, B., Yu, L., Li, X., Chen, M., Li, X., Hao, P., & Gong, P. (2021). A 1 km global cropland dataset from 10 000 BCE to 2100 CE. *Earth System Science Data*, 13(11), 5403–5421. <https://doi.org/10.5194/essd-13-5403-2021>
- Channell, R., & Lomolino, M. V. (2000). Trajectories to extinction: Spatial dynamics of the contraction of geographical ranges. *Journal of Biogeography*, 27(1), 169–179. <https://doi.org/10.1046/j.1365-2699.2000.00382.x>
- Chefaoui, R. M., & Lobo, J. M. (2008). Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological Modelling*, 210(4), 478–486. <https://doi.org/10.1016/j.ecolmodel.2007.08.010>
- Chen, G., Kéry, M., Plattner, M., Ma, K., & Gardner, B. (2012). Imperfect detection is the rule rather than the exception in plant distribution studies. *Journal of Ecology*, 9.
- Chen, M., Vernon, C. R., Graham, N. T., Hejazi, M., Huang, M., Cheng, Y., & Calvin, K. (2020). Global land use for 2015–2100 at 0.05° resolution under diverse socioeconomic and climate scenarios. *Scientific Data*, 7(1), 320. <https://doi.org/10.1038/s41597-020-00669-x>
- Colwell, R. K., & Rangel, T. F. (2009). Hutchinson's duality: The once and future niche. *Proceedings of the National Academy of Sciences of the United States of America*, 106(Suppl_2), 19651–19658. <https://doi.org/10.1073/pnas.0901650106>
- Dobrowski, S. Z., Thorne, J. H., Greenberg, J. A., Safford, H. D., Mynsberge, A. R., Crimmins, S. M., & Swanson, A. K. (2011). Modeling plant ranges over 75 years of climate change in California, USA: Temporal transferability and species traits. *Ecological Monographs*, 81(2), 241–257. <https://doi.org/10.1890/10-1325.1>
- Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, 17(1), 43–57. <https://doi.org/10.1111/j.1472-4642.2010.00725.x>
- Faurby, S., & Araújo, M. B. (2018). Anthropogenic range contractions bias species climate change forecasts. *Nature Climate Change*, 8(3), 252–256. <https://doi.org/10.1038/s41558-018-0089-x>
- Feeley, K. J., & Silman, M. R. (2011). The data void in modeling current and future distributions of tropical species: The paucity of tropical collections data. *Global Change Biology*, 17(1), 626–630. <https://doi.org/10.1111/j.1365-2486.2010.02239.x>
- Ficetola, G. F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M., Gielly, L., Lopes, C. M., Boyer, F., Pompanon, F., Rayé, G., & Taberlet, P. (2015). Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources*, 15(3), 543–556. <https://doi.org/10.1111/1755-0998.12338>
- Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12), 4302–4315. <https://doi.org/10.1002/joc.5086>
- Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24(1), 38–49. <https://doi.org/10.1017/S0376892997000088>
- Fourcade, Y., Besnard, A. G., & Secondi, J. (2018). Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Global Ecology and Biogeography*, 27(2), 245–256. <https://doi.org/10.1111/geb.12684>
- Fuller, R. A., McDonald-Madden, E., Wilson, K. A., Carwardine, J., Grantham, H. S., Watson, J. E., Klein, C. J., Green, D. C., & Possingham, H. P. (2010). Replacing underperforming protected areas achieves better conservation outcomes. *Nature*, 466(7304), 365. <https://doi.org/10.1038/nature09180>
- García, K., Lasco, R., Ines, A., Lyon, B., & Pulhin, F. (2013). Predicting geographic distribution and habitat suitability due to climate change of selected threatened forest tree species in the Philippines. *Applied Geography*, 44, 12–22. <https://doi.org/10.1016/j.apgeog.2013.07.005>
- GBIF.org. (2020, January 16). GBIF Occurrence Download. <https://doi.org/10.15468/dl.kvllgo>
- Gibson, L. M., Mychajliw, A. M., Leon, Y., Rupp, E., & Hadly, E. A. (2019). Using the past to contextualize anthropogenic impacts on the present and future distribution of an endemic Caribbean mammal. *Conservation Biology*, 33(3), 500–510. <https://doi.org/10.1111/cobi.13290>
- Girma, A., de Bie, C. A. J. M., Skidmore, A. K., Venus, V., & Bongers, F. (2016). Hyper-temporal SPOT-NDVI dataset parameterization captures species distributions. *International Journal of Geographical Information Science*, 30(1), 89–107. <https://doi.org/10.1080/13658816.2015.1082565>
- Gomes, V. H. F., Vieira, I. C. G., Salomão, R. P., & ter Steege, H. (2019). Amazonian tree species threatened by deforestation and climate change. *Nature Climate Change*, 9(7), 547–553. <https://doi.org/10.1038/s41558-019-0500-2>
- Goodwin, Z. A., Harris, D. J., Filer, D., Wood, J. R. I., & Scotland, R. W. (2015). Widespread mistaken identity in tropical plant collections. *Current Biology*, 25(22), R1066–R1067. <https://doi.org/10.1016/j.cub.2015.10.002>
- Guélat, J., & Kéry, M. (2018). Effects of spatial autocorrelation and imperfect detection on species distribution models. *Methods in Ecology and Evolution*, 9(6), 1614–1625. <https://doi.org/10.1111/2041-210X.12983>
- Guevara, L., Gerstner, B. E., Kass, J. M., & Anderson, R. P. (2018). Toward ecologically realistic predictions of species distributions: A cross-time example from tropical montane cloud forests. *Global Change Biology*, 24(4), 1511–1522. <https://doi.org/10.1111/gcb.13992>
- Guillera-Arroita, G. (2017). Modelling of species distributions, range dynamics and communities under imperfect detection: Advances, challenges and opportunities. *Ecography*, 40(2), 281–295. <https://doi.org/10.1111/ecog.02445>

- Guisan, A., Petitpierre, B., Broennimann, O., Daehler, C., & Kueffer, C. (2014). Unifying niche shift studies: Insights from biological invasions. *Trends in Ecology & Evolution*, 29(5), 260–269. <https://doi.org/10.1016/j.tree.2014.02.009>
- Guisan, A., & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. *Ecology Letters*, 8(9), 993–1009. <https://doi.org/10.1111/j.1461-0248.2005.00792.x>
- Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I. T., Regan, T. J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T. G., Rhodes, J. R., Maggini, R., Setterfield, S. A., Elith, J., Schwartz, M. W., Wintle, B. A., Broennimann, O., Austin, M., ... Buckley, Y. M. (2013). Predicting species distributions for conservation decisions. *Ecology Letters*, 16(12), 1424–1435. <https://doi.org/10.1111/ele.12189>
- Harisena, N. V., Groen, T. A., Toxopeus, A. G., & Naimi, B. (2021). When is variable importance estimation in species distribution modelling affected by spatial correlation? *Ecography*, 44(5), 778–788. <https://doi.org/10.1111/ecog.0553>
- Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS One*, 12(2), e0169748. <https://doi.org/10.1371/journal.pone.0169748>
- Hijmans, R. J., & Eten, J. V. (2012). Geographic analysis and modeling with raster data. *R Package Version*, 2, 1–25.
- Hijmans, R. J., Leathwick, J. R., & Elith, J. (2017). dismo: Species distribution modeling. *R Package Version*, 1(4), 1–1.
- Hirzel, A. H., Le Lay, G., Helfer, V., Randin, C., & Guisan, A. (2006). Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, 199(2), 142–152. <https://doi.org/10.1016/j.ecolmodel.2006.05.017>
- Holden, M. H., Yen, J. D. L., Briscoe, N. J., Lahoz-Monfort, J. J., Salguero-Gómez, R., Vesik, P. A., & Guillerá-Arroita, G. (2021). Assessing the accuracy of density-independent demographic models for predicting species ranges. *Ecography*, 44(3), 345–357. <https://doi.org/10.1111/ecog.05250>
- Hurtt, G. C., Chini, L., Sahajpal, R., Frolking, S., Bodirsky, B. L., Calvin, K., Doelman, J. C., Fisk, J., Fujimori, S., & Goldewijk, K. K. (2020). Harmonization of global land-use change and management for the period 850–2100 (LUH2) for CMIP6. *Geoscientific Model Development Discussions*, 1–65.
- Hutchinson, G. E. (1957). Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, 22(4), 415–427. <https://doi.org/10.1101/SQB.1957.022.01.039>
- Jarvie, S., & Svenning, J.-C. (2018). Using species distribution modeling to determine opportunities for trophic rewilding under future scenarios of climate change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1761), 20170446. <https://doi.org/10.1098/rstb.2017.0446>
- Leroy, B., Delsol, R., Hugué, B., Meynard, C. N., Barhoumi, C., Barbet-Massin, M., & Bellard, C. (2018). Without quality presence-absence data, discrimination metrics such as TSS can be misleading measures of model performance. *Journal of Biogeography*, 45(9), 1994–2002. <https://doi.org/10.1111/jbi.13402>
- Leroy, B., Meynard, C. N., Bellard, C., & Courchamp, F. (2016). Virtualspecies, an R package to generate virtual species distributions. *Ecography*, 39(6), 599–607. <https://doi.org/10.1111/ecog.01388>
- Liu, C., Berry, P. M., Dawson, T. P., & Pearson, R. G. (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, 28(3), 385–393. <https://doi.org/10.1111/j.0906-7590.2005.03957.x>
- Liu, C., Wolter, C., Xian, W., & Jeschke, J. M. (2020). Species distribution models have limited spatial transferability for invasive species. *Ecology Letters*, 23(11), 1682–1692. <https://doi.org/10.1111/ele.13577>
- Maiorano, L., Cheddadi, R., Zimmermann, N. E., Pellissier, L., Petitpierre, B., Pottier, J., Laborde, H., Hurdu, B. I., Pearman, P. B., Psomas, A., Singarayer, J. S., Broennimann, O., Vittoz, P., Dubuis, A., Edwards, M. E., Binney, H. A., & Guisan, A. (2013). Building the niche through time: Using 13,000 years of data to predict the effects of climate change on three tree species in Europe: Multi-temporal niche and species potential distribution. *Global Ecology and Biogeography*, 22(3), 302–317. <https://doi.org/10.1111/j.1466-8238.2012.00767.x>
- Manchego, C. E., Hildebrandt, P., Cueva, J., Espinosa, C. I., Stimm, B., & Günter, S. (2017). Climate change versus deforestation: Implications for tree species distribution in the dry forests of southern Ecuador. *PLoS One*, 13(4), e0190092. <https://doi.org/10.1371/journal.pone.0190092>
- Mangiacotti, M., Scali, S., Sacchi, R., Bassu, L., Nulchis, V., & Corti, C. (2013). Assessing the spatial scale effect of anthropogenic factors on species distribution. *PLoS One*, 8(6), e67573. <https://doi.org/10.1371/journal.pone.0067573>
- Marshall, L., Beckers, V., Vray, S., Rasmont, P., Vereecken, N. J., & Dendoncker, N. (2021). High thematic resolution land use change models refine biodiversity scenarios: A case study with Belgian bumblebees. *Journal of Biogeography*, 48(2), 345–358. <https://doi.org/10.1111/jbi.14000>
- Marshall, L., Biesmeijer, J. C., Rasmont, P., Vereecken, N. J., Dvorak, L., Fitzpatrick, U., Francis, F., Neumayer, J., Ødegaard, F., Paukkunen, J. P. T., Pawlikowski, T., Reemer, M., Roberts, S. P. M., Straka, J., Vray, S., & Dendoncker, N. (2018). The interplay of climate and land use change affects the distribution of EU bumblebees. *Global Change Biology*, 24(1), 101–116. <https://doi.org/10.1111/gcb.13867>
- Martínez-Freiria, F., Tarroso, P., Rebelo, H., & Brito, J. C. (2016). Contemporary niche contraction affects climate change predictions for elephants and giraffes. *Diversity and Distributions*, 22(4), 432–444. <https://doi.org/10.1111/ddi.12406>
- Maxwell, S. L., Fuller, R. A., Brooks, T. M., & Watson, J. E. M. (2016). Biodiversity: The ravages of guns, nets and bulldozers. *Nature* 536(7615), 143–145.
- McShea, W. J. (2014). What are the roles of species distribution models in conservation planning? *Environmental Conservation*, 41(02), 93–96. <https://doi.org/10.1017/S0376892913000581>
- Mertes, K., & Jetz, W. (2018). Disentangling scale dependencies in species environmental niches and distributions. *Ecography*, 41(10), 1604–1615. <https://doi.org/10.1111/ecog.02871>
- Meyer, C., Weigelt, P., & Kreft, H. (2016). Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecology Letters*, 19(8), 992–1006. <https://doi.org/10.1111/ele.12624>
- Meynard, C. N., Leroy, B., & Kaplan, D. M. Testing methods in species distribution modelling using virtual species: What have we learnt and what are we missing? *Ecography*, 42(12), 2021–2036. <https://doi.org/10.1111/ecog.04385>
- Milanesi, P., Della Rocca, F., & Robinson, R. A. (2020). Integrating dynamic environmental predictors and species occurrences: Toward true dynamic species distribution models. *Ecology and Evolution*, 10(2), 1087–1092. <https://doi.org/10.1002/ece3.5938>
- Newbold, T. (2018). Future effects of climate and land-use change on terrestrial vertebrate community diversity under different scenarios. *Proceedings of the Royal Society B: Biological Sciences*, 285(1881), 20180792. <https://doi.org/10.1098/rspb.2018.0792>
- Newbold, T., Hudson, L. N., Hill, S. L. L., Contu, S., Lysenko, I., Senior, R. A., Börger, L., Bennett, D. J., Choimes, A., Collen, B., Day, J., De Palma, A., Díaz, S., Echeverría-Londoño, S., Edgar, M. J., Feldman,

- A., Garon, M., Harrison, M. L. K., Alhousseini, T., ... Purvis, A. (2015). Global effects of land use on local terrestrial biodiversity. *Nature*, 520(7545), 45. <https://doi.org/10.1038/nature14324>
- Nisbet, R., Miner, G., & Yale, K. (2018). Chapter 11—Model evaluation and enhancement. In R. Nisbet, G. Miner, & K. Yale (Eds.), *Handbook of statistical analysis and data mining applications* (2nd ed., pp. 215–233). Academic Press. <https://doi.org/10.1016/B978-0-12-416632-5.00011-6>
- Norberg, A., Abrego, N., Blanchet, F. G., Adler, F. R., Anderson, B. J., Anttila, J., Araújo, M. B., Dallas, T., Dunson, D., Elith, J., Foster, S. D., Fox, R., Franklin, J., Godsoe, W., Guisan, A., O'Hara, B., Hill, N. A., Holt, R. D., Hui, F. K. C., ... Ovaskainen, O. (2019). A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. *Ecological Monographs*, 89(3), 1–24. <https://doi.org/10.1002/ecm.1370>
- Nüchel, J., Bøcher, P. K., Xiao, W., Zhu, A.-X., & Svenning, J.-C. (2018). Snub-nosed monkeys (*Rhinopithecus*): Potential distribution and its implication for conservation. *Biodiversity and Conservation*, 27(6), 1517–1538. <https://doi.org/10.1007/s10531-018-1507-0>
- Owens, H. L., Campbell, L. P., Dornak, L. L., Saupe, E. E., Barve, N., Soberón, J., Ingenloff, J., Lira-Noriega, A., Hensz, C. M., Myers, C. E., & Peterson, A. T. (2013). Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. *Ecological Modelling*, 263, 10–18. <https://doi.org/10.1016/j.ecolmodel.2013.04.011>
- Pang, S. E. H., De Alban, J. D. T., & Webb, E. L. (2021). Effects of climate change and land cover on the distributions of a critical tree family in the Philippines. *Scientific Reports*, 11(1), 276. <https://doi.org/10.1038/s41598-020-79491-9>
- Peterson, A. T., Cobos, M. E., & Jiménez-García, D. (2018). Major challenges for correlational ecological niche model projections to future climate conditions: Climate change, ecological niche models, and uncertainty. *Annals of the New York Academy of Sciences*, 1429(1), 66–77. <https://doi.org/10.1111/nyas.13873>
- Petitpierre, B., Kueffer, C., Broennimann, O., Randin, C., Daehler, C., & Guisan, A. (2012). Climatic niche shifts are rare among terrestrial plant invaders. *Science*, 335(6074), 1344–1348. <https://doi.org/10.1126/science.1215933>
- Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E., & Blair, M. E. (2017). Opening the black box: An open-source release of Maxent. *Ecography*, 40(7), 887–893. <https://doi.org/10.1111/ecog.03049>
- Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3–4), 231–259. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>
- Prestele, R., Alexander, P., Rounsevell, M. D. A., Arneth, A., Calvin, K., Doelman, J., Eitelberg, D. A., Engström, K., Fujimori, S., Hasegawa, T., Havlik, P., Humpenöder, F., Jain, A. K., Krisztin, T., Kyle, P., Meiyappan, P., Popp, A., Sands, R. D., Schaldach, R., ... Verburg, P. H. (2016). Hotspots of uncertainty in land-use and land-cover change projections: A global-scale model comparison. *Global Change Biology*, 22(12), 3967–3983. <https://doi.org/10.1111/gcb.13337>
- R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 55, 275–286.
- Requena-Mullor, J. M., Maguire, K. C., Shinneman, D. J., & Caughlin, T. T. (2019). Integrating anthropogenic factors into regional-scale species distribution models—A novel application in the imperiled sagebrush biome. *Global Change Biology*, 25(11), 3844–3858. <https://doi.org/10.1111/gcb.14728>
- Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Aroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8), 913–929. <https://doi.org/10.1111/ecog.02881>
- Rutrough, A., Widick, I. V., & Bean, W. T. (2019). Reconstruction of the historical range alters niche estimates in an endangered rodent. *Ecography*, 42(10), 1742–1751. <https://doi.org/10.1111/ecog.04238>
- Ryo, M., Aguilar-Trigueros, C. A., Pinek, L., Muller, L. A. H., & Rillig, M. C. (2019). Basic principles of temporal dynamics. *Trends in Ecology & Evolution*, 34(8), 723–733. <https://doi.org/10.1016/j.tree.2019.03.007>
- Saah, D., Tenneson, K., Poortinga, A., Nguyen, Q., Chishtie, F., Aung, K. S., Markert, K. N., Clinton, N., Anderson, E. R., Cutter, P., Goldstein, J., Housman, I. W., Bhandari, B., Potapov, P. V., Matin, M., Uddin, K., Pham, H. N., Khanal, N., Maharjan, S., ... Ganz, D. (2020). Primitives as building blocks for constructing land cover maps. *International Journal of Applied Earth Observation and Geoinformation*, 85, 1–18. <https://doi.org/10.1016/j.jag.2019.101979>
- Saupe, E. E., Barve, V., Myers, C. E., Soberón, J., Barve, N., Hensz, C. M., Peterson, A. T., Owens, H. L., & Lira-Noriega, A. (2012). Variation in niche and distribution model performance: The need for a priori assessment of key causal factors. *Ecological Modelling*, 237–238, 11–22. <https://doi.org/10.1016/j.ecolmodel.2012.04.001>
- Scheele, B. C., Foster, C. N., Banks, S. C., & Lindenmayer, D. B. (2017). Niche contractions in declining species: Mechanisms and consequences. *Trends in Ecology & Evolution*, 32(5), 346–355. <https://doi.org/10.1016/j.tree.2017.02.013>
- Seaborn, T., Goldberg, C. S., & Crespi, E. J. (2021). Drivers of distributions and niches of North American cold-adapted amphibians: Evaluating both climate and land use. *Ecological Applications*, 31(2). <https://doi.org/10.1002/eap.2236>
- Silva, D. E., Badeau, V., Legay, M., Corcket, E., & Dupouey, J.-L. (2012). Tracking human impact on current tree species distribution using plant communities. *Journal of Vegetation Science*, 23(2), 313–324. <https://doi.org/10.1111/j.1654-1103.2011.01341.x>
- Smith, A. B., & Santos, M. J. (2020). Testing the ability of species distribution models to infer variable importance. *Ecography*, 43(12), 1801–1813. <https://doi.org/10.1111/ecog.05317>
- Soberón, J., & Peterson, A. T. (2005). Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics*, 2, 1–10. <https://doi.org/10.17161/bi.v2i0.4>
- Soley-Guardia, M., Gutiérrez, E. E., Thomas, D. M., Ochoa-G, J., Aguilera, M., & Anderson, R. P. (2016). Are we overestimating the niche? Removing marginal localities helps ecological niche models detect environmental barriers. *Ecology and Evolution*, 6(5), 1267–1279. <https://doi.org/10.1002/ece3.1900>
- Struebig, M. J., Wilting, A., Gaveau, D. L. A., Meijaard, E., Smith, R. J., Fischer, M., Metcalfe, K., Kramer-Schadt, S., Abdullah, T., Abram, N., Alfred, R., Ancrenaz, M., Augeri, D. M., Belant, J. L., Bernard, H., Bezuijen, M., Boonman, A., Boonratana, R., Boorsma, T., ... Wong, A. (2015). Targeted conservation to safeguard a biodiversity hotspot from climate and land-cover change. *Current Biology*, 25(3), 372–378. <https://doi.org/10.1016/j.cub.2014.11.067>
- Titeux, N., Henle, K., Mihoub, J.-B., Regos, A., Geijzendorffer, I. R., Cramer, W., Verburg, P. H., & Brotons, L. (2016). Biodiversity scenarios neglect future land-use changes. *Global Change Biology*, 22(7), 2505–2515. <https://doi.org/10.1111/gcb.13272>
- 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? *Ecological Modelling*, 220(4), 589–594. <https://doi.org/10.1016/j.ecolmodel.2008.11.010>
- Verburg, P. H., Neumann, K., & Nol, L. (2011). Challenges in using land use and land cover data for global change studies. *Global Change Biology*, 17(2), 974–989. <https://doi.org/10.1111/j.1365-2486.2010.02307.x>

- Vignali, S., Barras, A., & Braunisch, V. (2019). SDMtune: Species distribution model selection. *R Package Version 1.0.1*. <https://github.com/ConsBiol-unibern/SDMtune>
- Warren, D. L., Glor, R. E., & Turelli, M. (2008). Environmental niche equivalency versus conservatism: Quantitative approaches to niche evolution. *Evolution*, 62(11), 2868–2883. <https://doi.org/10.1111/j.1558-5646.2008.00482.x>
- Warren, D. L., Matzke, N. J., & Iglesias, T. L. (2019). Evaluating species distribution models with discrimination accuracy is uninformative for many applications. *Ecology*, 47(1), 167–180. <https://doi.org/10.1111/jbi.13705>
- Warren, D. L., Matzke, N. J., & Iglesias, T. L. (2020). Evaluating presence-only species distribution models with discrimination accuracy is uninformative for many applications. *Journal of Biogeography*, 47(1), 167–180. <https://doi.org/10.1111/jbi.13705>
- Winkler, K., Fuchs, R., Rounsevell, M., & Herold, M. (2021). Global land use changes are four times greater than previously estimated. *Nature Communications*, 12(1), 2501. <https://doi.org/10.1038/s41467-021-22702-2>
- Yates, K. L., Bouchet, P. J., Caley, M. J., Mengersen, K., Randin, C. F., Parnell, S., Fielding, A. H., Bamford, A. J., Ban, S., Barbosa, A. M., Dormann, C. F., Elith, J., Embling, C. B., Ervin, G. N., Fisher, R., Gould, S., Graf, R. F., Gregr, E. J., Halpin, P. N., ... Sequeira, A. M. M. (2018). Outstanding challenges in the transferability of ecological models. *Trends in Ecology & Evolution*, 33(10), 790–802. <https://doi.org/10.1016/j.tree.2018.08.001>
- Zhu, G., Li, H., & Zhao, L. (2017). Incorporating anthropogenic variables into ecological niche modeling to predict areas of invasion of *Popillia japonica*. *Journal of Pest Science*, 90(1), 151–160. <https://doi.org/10.1007/s10340-016-0780-5>
- Zurell, D., Berger, U., Cabral, J. S., Jeltsch, F., Meynard, C. N., Münkemüller, T., Nehrbass, N., Pagel, J., Reineking, B., Schröder, B., & Grimm, V. (2010). The virtual ecologist approach: Simulating data and observers. *Oikos*, 119(4), 622–635. <https://doi.org/10.1111/j.1600-0706.2009.18284.x>

BIOSKETCH

Sean Pang is a spatial ecologist and currently a PhD candidate at the National University of Singapore. His research interests revolve around developing robust spatial models to better understand the impacts of climate change and land cover change on species distributions, and how these models can be applied to conservation and management. Sean is also interested in advancing the methods and analyses surrounding species distribution models. Twitter at https://twitter.com/SEH_Pang

Author contributions: S.E.H.P., Y.W., J.D.T.D.A. and E.L.W. contributed to the planning of this study and interpretation of the data. S.E.H.P. designed and carried out the simulations, modelling and data analyses. S.E.H.P. and E.L.W. wrote the first draft of the manuscript. All authors provided feedback on the manuscript and approved its final version.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Pang, S. E. H., Zeng, Y., De Alban, J. D. T., & Webb, E. L. (2022). Occurrence–habitat mismatching and niche truncation when modelling distributions affected by anthropogenic range contractions. *Diversity and Distributions*, 28, 1327–1343. <https://doi.org/10.1111/ddi.13544>