

## Research

### The effect of positional error on fine scale species distribution models increases for specialist species

Lukáš Gábor, Vítězslav Moudrý, Vincent Lecours, Marco Malavasi, Vojtěch Barták, Michal Fogl, Petra Šimová, Duccio Rocchini and Tomáš Václavík

L. Gábor (<https://orcid.org/0000-0001-6137-0994>), V. Moudrý (<https://orcid.org/0000-0002-3194-451X>) ✉ ([moudry@fzp.czu.cz](mailto:moudry@fzp.czu.cz)), M. Malavasi, V. Barták (<https://orcid.org/0000-0001-9887-1290>), M. Fogl (<https://orcid.org/0000-0002-5880-6926>), P. Šimová (<https://orcid.org/0000-0003-2480-1171>) and D. Rocchini (<https://orcid.org/0000-0003-0087-0594>), Dept of Applied Geoinformatics and Spatial Planning, Faculty of Environmental Sciences, Czech Univ. of Life Sciences Prague, Praha – Suchbátka, Czech Republic. DR also at: Univ. of Trento, Center Agriculture Food Environment (C3A), S. Michele all'Adige, TN, Italy, and Univ. of Trento, Dept of Cellular, Computational and Integrative Biology – CIBIO, Univ. of Trento, Povo, Italy, and Fondazione Edmund Mach, Research and Innovation Centre, Dept of Biodiversity and Molecular Ecology, S. Michele all'Adige, TN, Italy. – V. Lecours (<https://orcid.org/0000-0002-4777-3348>), School of Forest Resources and Conservation, Univ. of Florida, Gainesville, FL, USA. – T. Václavík (<https://orcid.org/0000-0002-1113-6320>), Palacký Univ. Olomouc, Dept of Ecology and Environmental Sciences, Faculty of Science, Olomouc, Czech Republic, and UFZ – Helmholtz Centre for Environmental Research, Dept of Computational Landscape Ecology, Leipzig, Germany.

#### Ecography

42: 1–14, 2019

doi: 10.1111/ecog.04687

Subject Editor: Dan Warren  
Editor-in-Chief: Miguel Araújo  
Accepted 4 October 2019



Species occurrences inherently include positional error. Such error can be problematic for species distribution models (SDMs), especially those based on fine-resolution environmental data. It has been suggested that there could be a link between the influence of positional error and the width of the species ecological niche. Although positional errors in species occurrence data may imply serious limitations, especially for modelling species with narrow ecological niche, it has never been thoroughly explored. We used a virtual species approach to assess the effects of the positional error on fine-scale SDMs for species with environmental niches of different widths. We simulated three virtual species with varying niche breadth, from specialist to generalist. The true distribution of these virtual species was then altered by introducing different levels of positional error (from 5 to 500 m). We built generalized linear models and MaxEnt models using the distribution of the three virtual species (unaltered and altered) and a combination of environmental data at 5 m resolution. The models' performance and niche overlap were compared to assess the effect of positional error with varying niche breadth in the geographical and environmental space. The positional error negatively impacted performance and niche overlap metrics. The amplitude of the influence of positional error depended on the species niche, with models for specialist species being more affected than those for generalist species. The positional error had the same effect on both modelling techniques. Finally, increasing sample size did not mitigate the negative influence of positional error. We showed that fine-scale SDMs are considerably affected by positional error, even when such error is low. Therefore, where new surveys are undertaken, we recommend paying attention to data collection techniques to minimize the positional error in occurrence data and thus to avoid its negative effect on SDMs, especially when studying specialist species.

Keywords: data errors, niche breadth, spatial overlay, virtual species



[www.ecography.org](http://www.ecography.org)

© 2019 The Authors. Ecography published by John Wiley & Sons Ltd on behalf of Nordic Society Oikos  
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

## Introduction

Studying relationships between species and their environment is fundamental for understanding Earth's biodiversity. Species distribution models (SDMs) are a common tool used to study these relationships. They use species occurrence data and environmental data to produce a set of rules explaining the environmental space where species were collected or observed (Ferrier et al. 2017). All applications of SDMs, however, assume that species occurrence data are largely free of spatial error. Nonetheless, all spatial data inherently contain some level and type of spatial errors. These errors can be, for example, related to the use of inadequate spatial resolution (Gottschalk et al. 2011, Šímová et al. 2019), low sample size (Wisz et al. 2008, Moudrý et al. 2017), biased sampling (Hijmans 2012, Ranc et al. 2016) or occurrences with positional error (Graham et al. 2008, Osborne and Leitão 2009, Mitchell et al. 2017). Data quality (both for species occurrences and environmental variables) is currently considered a major factor limiting SDM accuracy (Araújo et al. 2019) and demonstrating, quantifying and understanding the consequences of these errors is therefore critical.

It is often assumed that the negative effects of positional error (i.e. inaccurate location of species occurrences) is minimal or mainly associated with relatively older datasets that are often georeferenced from textual descriptions of their locations (which may cause errors of up to hundreds of meters, Wieczorek et al. 2004). However, it is also necessary to consider positional errors inherent to data georeferenced using modern global navigation satellite systems (GNSS). The positional error of GNSS data may be caused by the use of outdated technology, by poor satellite signal reception (e.g. because of inappropriate site conditions), or by data processing (e.g. conversion between coordinate systems or rounding of coordinate values). Moreover, species occurrence data often represent the position of the observer and not the actual position of the species (Zhang et al. 2018). Additionally, where the marine environment is concerned, species data are often acquired using underwater cameras, in which case the positional error can be affected for example by the camera depth; the deeper the camera is, the greater is the positional error (Ratray et al. 2014, Mitchell et al. 2017). Therefore, even though the accuracy of standard GNSS is usually below 30 m (Frair et al. 2010), the errors associated with such data may be much larger.

In addition, performance of SDMs is complicated by various spatial (e.g. prevalence or range size) and ecological (e.g. niche breadth) characteristics of the studied species (Luoto et al. 2005, Bulluck et al. 2006, McPherson and Jetz 2007, Evangelista et al. 2008, Chefaoui et al. 2011, Connor et al. 2018). It has been hypothesized that range size is positively correlated with niche breadth (i.e. the range of environments that the species can inhabit), in other words that species able to tolerate a wider range of conditions are typically more widespread (Brown 1984, Gaston et al. 1997, Arribas et al. 2012, Boulangeat et al. 2012). The niche

breadth–range size relationship is one of the possible mechanisms explaining commonness and rarity. Modelling rare species (i.e. species with small geographical ranges) is particularly problematic and novel approaches have been adopted for this purpose (Breiner et al. 2015) to overcome the common problem of a low number of occurrences available for modelling that may not be sufficient to completely describe the species niche. Similar effects can be caused by a low positional accuracy of the occurrences (Johnson and Gillingham 2008, Fernandez et al. 2009, Osborne and Leitão 2009).

Although the magnitude of the niche breadth–range size relationship is still under debate, a recent meta-analysis of 64 studies found a significant positive relationship between the range size and niche breadth (Slatyer et al. 2013). Such a synergic relationship can increase the already high vulnerability of specialist species to environmental changes. In addition, Slatyer et al. (2013) suggested that specialist species might be particularly vulnerable to any environmental change due to synergistic effects of a narrow niche and small range size. Specialist species are of high conservation concern, and SDMs might be the only tractable means of estimating their distribution and reaction to environmental change. However, confounding effects of inaccurate data on modelling species that utilize a narrow niche breadth (i.e. specialist) versus species that utilize a wide niche breadth (i.e. generalist) are unknown (Connor et al. 2018).

It is intuitive that positional error of a given magnitude might have a greater effect on specialist than generalist species, as it is more likely that occurrences get incorrectly shifted into cells representing an unsuitable environment, i.e. environment that is outside of the species' environmental niche. This, however, has never been thoroughly explored because it is extremely difficult, if not impossible, to estimate the true responses of a real species to the environment and, consequently, to be able to fully understand the true suitability of an area for the species in question.

In this study, we focused on Light Detection and Ranging (LiDAR)-derived variables that are being more and more often combined with species distribution data of unknown positional accuracy to study species–environment relationships at fine scales. Studies published so far have used real species to test the effect of positional error. However, real species distribution data are usually affected by a complex set of other uncertainties (e.g. sampling bias, incompleteness, inaccuracies). As a consequence, the isolation and identification of the effects of positional error can be very challenging, if not impossible. This is likely one of the reasons why little consensus exists on how the effect of positional error manifests in SDMs (Naimi et al. 2011, Mitchell et al. 2017). For example, Graham et al. (2008) concluded that SDMs are robust to positional error while others argued that positional errors reduce models' performance (Johnson and Gillingham 2008, Fernandez et al. 2009, Osborne and Leitão 2009).

Another aspect may be that positional errors of species occurrences were studied using relatively coarse environmental data (but see Mitchell et al. 2017). Positional error

Table 1. Overview of prior studies focused on the influence of positional error in species occurrence data on SDMs.

	Species data	Environmental data	Resolution of input environmental data (pixel size)	Range of shifting occurrences	
Graham et al. 2008	observed	categorical, continuous	100×100m	0–5 km	0–50 pixels
Johnson and Gillingham 2008	observed	categorical	30×30 m	50–1000m (over 50 m)	1–34 pixels
Osborne and Leitão 2009	observed	continuous	1×1 km	0–1, 2–3, 4–5, 0–5 km	0–1, 2–3, 4–5, 0–5 pixels
Fernandez et al. 2009	observed	continuous	1×1 km	5–10–25–50 km	1–5, 1–10, 1–25, 1–50 pixels
Naimi et al. 2011	artificial	continuous	artificial data	x	1–30 (over 1 pixel)
Mitchell et al. 2017	observed	continuous	2.5×2.5 m	5–25–50–20–400 m	1–2, 1–12, 1–80, 1–160 pixels

considered in prior studies ranged from 50 m up to 50 km (Table 1). While such error results in a shift over several cells in a coarse-resolution SDM (e.g. 1×1 km), it will cause a much greater shift in a fine-resolution SDM (e.g. 10×10 m). Therefore, with the increasing availability of fine-scale data, additional studies are needed (Osborne and Leitão 2009); it can be expected that SDMs at fine scales would be more sensitive to positional error.

To ensure the full knowledge of the exact ecological and geographical characteristics of the species and to avoid unknown complexities associated with real data, we used a virtual species approach to test the effect of the positional error in species occurrences on fine-scale SDMs in the context of species niche breadth (i.e. specialist versus generalist species). We generated three virtual species that differed in characteristics related to the geographic distribution of the species, i.e. prevalence and relative occurrence area (ROA); the proportion of the total study area occupied by the species (Lobo 2008).

The virtual species approach allowed us to control the experiment and to isolate the effects of positional error (Zurell et al. 2010). This approach is increasingly used to evaluate the effects of data inaccuracies on model performance (Barbet-Massin et al. 2012, Václavík and Meentemeyer 2012, Qiao et al. 2015, Ranc et al. 2016, Fernandes et al. 2018, Leroy et al. 2018, Moudrý et al. 2018, Gábor et al. 2019, Meynard et al. 2019), but has yet to be adopted for the study of positional error. In particular, we tested whether: 1) SDMs for specialist species are more affected by positional error than those for generalist species; 2) it is possible to compensate the assumed negative effect of a positional error with a higher sample size; and 3) the positional error has different effects when using a parametric (e.g. generalized linear model) versus a nonparametric (e.g. MaxEnt) modelling technique.

## Material and methods

### LiDAR data acquisition, processing and variable selection

Discrete LiDAR data were collected in Krkonose Mountains National Park (KRNAP), Czech Republic (Supplementary material Appendix 1 Fig. A1) in 2012 using a small-footprint airborne LiDAR system (RIEGL LMS Q-680i). The average point density was approximately six points per square meter. The LiDAR point cloud was automatically classified into ground, vegetation, building, wire and transmission tower classes in the ENVI LiDAR software (ver. 5.3) and LAsTools (ver. 171215). The terrain data points were used to produce a digital terrain model (DTM), and the vegetation data points were used to produce a canopy height model (CHM) (Khosravipour et al. 2016). Both models were generated from the point cloud at a 0.5 m resolution and subsequently resampled to 5 m cell resolution for the analysis to improve processing time. A topographic wetness index (TWI) was derived from the DTM based on the equation

$$TWI = \ln \left( \frac{A_s}{\tan \beta} \right)$$

where  $A_s$  is the specific catchment area and  $\tan \beta$  is the local slope in radians (Beven and Kirkby 1979). To calculate the specific catchment area, we used the multiple flow routing algorithm of Quinn et al. (1991), recommended by Kopecký and Čížková (2010), using SAGA-GIS (Conrad 2003).

The selection of these three variables (DTM, CHM, TWI) was motivated by the need to simulate a realistic situation that includes variables with various levels of spatial

autocorrelation (Supplementary material Appendix 2 Fig. A2). CHM describes a horizontal structural variability of the vegetation and is known to affect species richness (Lefsky et al. 2002). For example, higher vegetation was found to be related to higher bird species richness (Davies and Asner 2014). TWI is a surrogate for soil moisture, an environmental variable that affects the vegetation composition and that has been previously used to predict bird occurrences (Besnard et al. 2013, Reif et al. 2018). The relationships between CHM and TWI on the one side and bird distribution and richness on the other side make our study relatable to applications with real species; our virtual species could theoretically be birds with specific habitat requirements in terms of terrain characteristic and vegetation structure. We also used the DTM as a surrogate for climatic variables and to restrict our virtual species to certain altitudes (Coops et al. 2010, Vogeler et al. 2014).

### Simulating virtual species with different niche breadths

Virtual species were generated with the `virtualespecies` package (Leroy et al. 2016) in the statistical software R v.3.4.4 (R Development Core Team). The process involved three steps: a) generating the true distribution of the virtual species' environmental suitability, b) converting the environmental suitability into presences and absences and c) sampling species occurrences for further analysis and modelling.

Applying the `formatFunctions` function in R, we defined the species–environment relationships using normal distribution curves. To simulate species with different niche breadth, prevalence and ROA, we used the same means and varied standard deviations of the used environmental variables (Supplementary material Appendix 3 Table B1). Specifically, we simulated three distinct virtual species with varying ROAs and prevalence that represent realistic scenarios of species' extent of occurrence in the study area. The species with low ROA (4%) represents a specialist with low species prevalence (0.04), narrow niche breadth and small geographical range. The species with medium ROA (12%) may be described as an intermediate species (species prevalence = 0.12) with a wider niche breadth and medium geographical range. Finally, the species with high ROA (52%) can be perceived as a generalist with high species prevalence (0.47), wide niche breadth and wide geographical range (Futuyma and Moreno 1988, Devictor et al. 2010, Franklin 2010, Peers et al. 2012). Subsequently, we multiplied individual species' responses to environmental variables in order to acquire an environmental suitability raster (function `generateSpFromFun`). We opted for multiplication of the variables to assume irreplaceability of environmental conditions (i.e. we assumed that unsuitability of one condition causes a low probability of occurrence even though remaining conditions are in species' range of suitable values).

As noted in several studies (Meynard and Kaplan 2012, 2013, Moudrý 2015, Meynard et al. 2019), an appropriate setting of the whole simulation with respect to the research questions is crucial for obtaining reliable results. In addition,

Meynard et al. (2019) highlighted that simulation studies based on the threshold approach fail in appropriately separating factors such as prevalence and niche breadth. Therefore, due to these concerns, we adopted a probabilistic simulation approach (logistic function with  $\alpha = -0.05$  and  $\beta = 0.3$ ) to convert the environmental suitability rasters into probabilities of occurrences that were subsequently used to sample binary presence/absence rasters (function `convertToPA`). To sample species occurrences (function `sampleOccurrences`), we randomly generated, using a uniform random distribution, both presence-only and presence/absence data. Both types of occurrence datasets were generated in order to test different modelling techniques (cf. section Model fitting and evaluation). To test whether it is possible to compensate the assumed negative effect of positional error with a higher sample size, we generated four different sample sizes. Specifically, 30, 100, 500 and 1000 species presences were generated, complemented for the purpose of GLM modelling by twice as many absences.

### Simulating positional error in species occurrences

It is generally assumed that the magnitude of the positional error in species occurrence varies based on the source of the error. The positional error associated with GNSS points (e.g. species occurrences) may range from a few centimetres up to several metres. Furthermore, in some species such as birds or big predators, it is usually impossible to record their accurate position and such data are shifted by tens or hundreds of meters. An even greater shift is sometimes observed in museum databases. Therefore, to evaluate the range of possible magnitudes of the positional error, we simulated the positional error by shifting the sampled locations (i.e. presences and, in case of GLM, also absences) in a random direction according to six scenarios that corresponded to different distances ranging from 5–10 m up to 100–500 m. The error in the focal virtual species locations was 5–10 m for S1 scenario, 10–15 m for S2, 15–20 m for S3, 20–50 m for S4, 50–100 m for S5 and 100–500 m for S6 (Supplementary material Appendix 4 Table C1). Scenarios S1–S4 simulated realistic degrees of error if using modern monitoring technologies like GNSS, while scenarios S5–S6 simulated more extreme positional errors that could be associated with species observations recorded without GNSS, species difficult to pinpoint properly such as birds or big predators, or occurrences from museum databases. If the shifting of the original data points resulted in the points falling outside the study area, we recalculated the shift until the new coordinates were located within the boundaries of the study area. We provide a script of how we simulated virtual species and shifting occurrences in Supplementary material Appendix 2.

### Model fitting and evaluation

We selected generalized linear models (GLM; Nelder and Baker 1972, Oksanen and Minchin 2002) as a presence/absence method and MaxEnt (Phillips et al. 2006) as a

presence-background method that are often adopted in ecological studies (Moudrý and Šimová 2013, Linda et al. 2016, Malavasi et al. 2018, Gábor et al. 2019, Watts et al. 2019). In addition, Graham et al. (2008) showed that these two approaches were among the better performing modelling techniques when the data was affected by positional errors. Models were built in the statistical software R using the ‘dismo’ (ver. 1.1.4) and ‘glm2’ (ver. 1.2.1) packages. The GLM was run with a logit–link function and binomial distribution. The quadratic terms of the three environmental variables were included because of the known normal distribution curves of the response function. To enable the comparison of individual SDMs, we needed to maintain the parameters of MaxEnt unchanged, as done in many prior studies (Franklin et al. 2014, Fourcade et al. 2014, Holloway et al. 2016, Ranc et al. 2016, Tingley et al. 2018, Ye et al. 2018). The default settings established by Phillips et al. (2009) were used with randomly drawn background data generated from the binary map of the true occurrences of the virtual species. The same three environmental variables (DTM, CHM and TWI) used in the process of generating virtual species were used in the SDMs. Fivefold cross-validation where the data were randomly divided into fifths was used to evaluate the models. Four fifths of the data were used to train the model and the remaining one fifth was used to assess the performance. Control models without positional error were calculated for all three species with different niche breadth, prevalence and ROA and for both modelling techniques, allowing an easy comparison of the effect of positional error on model performance.

The area under the receiver operating characteristic curve (AUC) (Fielding and Bell 1997, Jiménez-Valverde 2012) and the true-skill statistic (TSS) (Allouche et al. 2006) were used to assess model performance (i.e. discrimination accuracy). AUC is widely used in ecological studies as a single threshold-independent measure of model performance (Václavík and Meentemeyer 2012, Mitchell et al. 2017). The AUC ranges from 0 to 1 where a score of 1 indicates perfect discrimination, a score of 0.5 indicates random performance and values lower than 0.5 indicate a worse than random performance. TSS is a frequently used threshold dependent metric (Cianfrani et al. 2018, Eaton et al. 2018) taking both omission and commission errors into account. It ranges from  $-1$  to  $+1$  where  $+1$  indicates perfect agreement and values of zero or less indicate random performance (Allouche et al. 2006).

To quantify differences between the true probability of occurrence of virtual species and the predicted distribution inferred from the models in geographical space, their niche overlap was compared using the  $I$  measure (Warren et al. 2008, Rödder and Engler 2011) and Spearman’s rank correlation. The  $I$  ranges between 0 (no overlap) and 1 (perfect overlap). Following Rödder and Engler (2011), we used the following classes to interpret the results: no or very limited overlap (0–0.2), low overlap (0.2–0.4), moderate overlap (0.4–0.6), high overlap (0.6–0.8) and very high overlap (0.8–1.0). Spearman’s rank correlation ranges between  $-1$  and  $+1$ , where  $-1$  indicates that species responses to the environment

are exactly negatively correlated (opposite) and  $+1$  indicates perfectly positively correlated overlap (identical). The closer the values are to zero, the lower is the niche overlap.

The magnitude of the negative effect of the positional error on SDMs is dependent on the size of the positional error and distribution of species’ suitable environment in the geographical space (Naimi et al. 2011). The positional data may be shifted in the geographical space and even a relatively low positional error in geographical space can have a profound effect on environmental niche estimates in environmental space and vice versa. Furthermore, we expected this would be related to the species niche breadth. Therefore, we were also interested in how the positional error is manifested in the environmental space and measured the niche overlap in the environmental space as well. We used  $I$  and Spearman’s rank correlation implemented in ENMTools 0.2 (Warren et al. 2019a, b) to estimate overlap in the environmental space between models fitted with accurate occurrences without any positional error (hereafter unaltered models) and models fitted with shifted occurrences (i.e. scenarios S1–S6).

We ran the entire process from species generation to model evaluation 30 times (Fig. 1). In addition, we used the analysis of variance (ANOVA) to assess the strength of the individual effects of the positional error, sample size, ROA and modelling technique, including all possible interactions. We compared the relative importance of individual predictors based on their contribution to the overall explained variation ( $R^2$ ). Instead of formal testing, we plotted the effects (and their confidence intervals) of all predictors combinations and evaluated them qualitatively. Because both AUC and TSS values were highly heteroscedastic (e.g. the ratio between maximum and minimum standard deviation across all factors combinations was 22 resp. 19 for AUC resp. TSS), we used robust variance–covariance matrix estimator suggested by MacKinnon and White (1985) for computation of confidence intervals. This was done using an R package ‘sandwich’ (Zeileis 2006).

## Results

### Unaltered models

Both performance metrics (AUC and TSS) largely followed the same pattern and highlighted excellent model performance for all, i.e. specialist, intermediate and generalist, species (AUC ranged from 0.91 up to 0.97 for MaxEnt models and from 0.80 up to 0.85 for GLM models). The only exception were the MaxEnt models for generalist species where AUC achieved only good performance (mean AUC 0.73). MaxEnt models were more successful in modelling specialist and intermediate species while GLM models were more accurate for the generalist species (Fig. 2).

Models achieved high or very high niche overlaps in geographical space according to both  $I$  and Spearman’s rank correlation. In general, the niche overlap decreased in the following order: generalist, specialists and intermediate species,

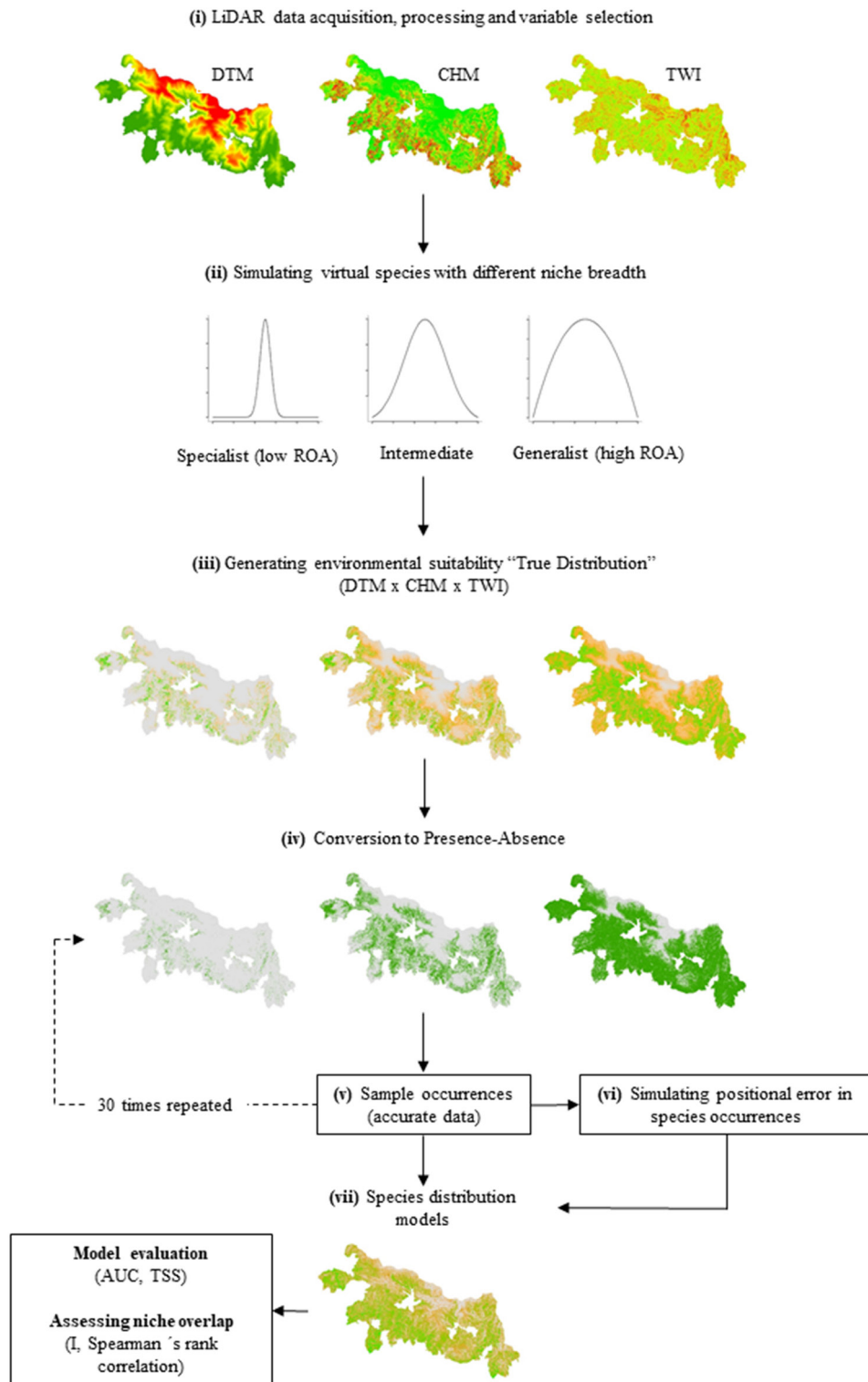


Figure 1. General modelling process. (i) We first acquired and processed LiDAR data and selected three fine-scale environmental predictors: DTM, CHM and TWI. (ii) We simulated virtual species with different niche breadths (ROA) by defining their response to environmental gradients for each environmental variable. (iii) We multiplied those variables to generate environmental suitability ('true' distribution of virtual species). (iv) We translated the probability of species occurrence to a presence-absence raster. (v) We sampled occurrences based on the presence-absence raster. (vi) We simulated the positional error in species occurrences. (vii) We generated SDMs with accurate as well as shifted occurrences, evaluated their performances (AUC, TSS) and assessed the niche overlap (I, Spearman's rank correlation) in the geographical and environmental space.

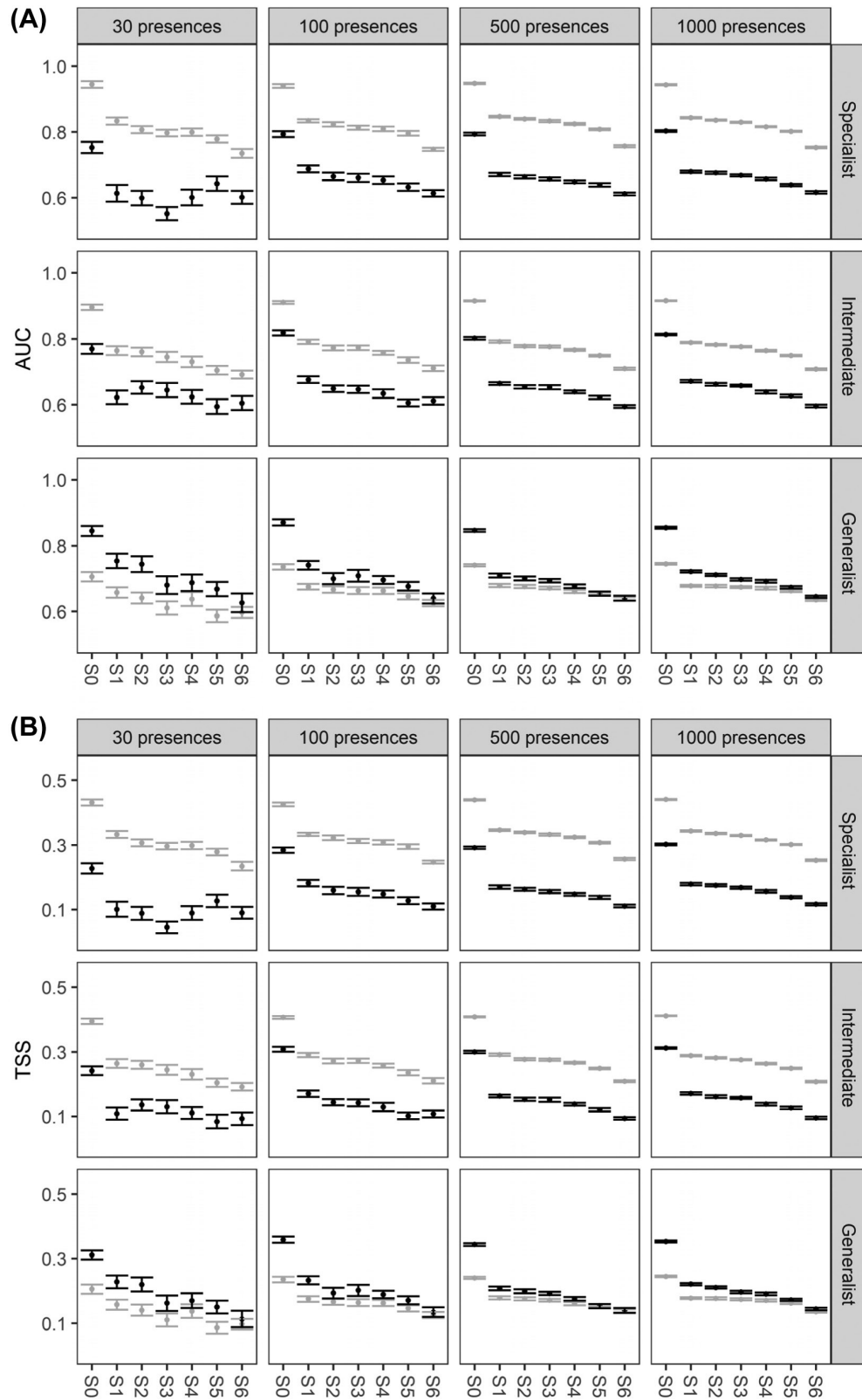


Figure 2. Resulting AUC (A) and TSS (B) scores according to different species niche breadth (specialist, intermediate, generalist), positional error (S0, unaltered models; S1, 5–10 m; S2, 10–15 m; S3, 15–20 m; S4 20–50 m, S5, 50–100 m; S6, 100–500 m) and sample size (number of presences = 30, 100, 500, 1000; note that for GLM models twice as many absences compared to presences were generated). Black colour shows results for GLM models while grey shows results for MaxEnt models.

except for the Spearman's rank correlation for specialists modelled by MaxEnt that achieved very high correlation. Comparison of modelling techniques showed that MaxEnt models achieved a higher niche overlap than GLM for all species with the most obvious differences in specialist species. An increase in the sample size of unaltered models led to none or negligible increase in niche overlap (Fig. 3).

### Effect of positional error on models of species with different niche breadth

Results show, independently of the modelling technique, a clear trend of the positional error worsening model performance (both AUC and TSS). The highest drop is evident between unaltered models and models affected by the smallest simulated positional error (5–10 m). Increasing the positional error further led to additional decrease in model performances; however, this decrease was minimal (positional error 10–50 m). Even the extreme cases of positional error (50–100 and 100–500 m) led to a relatively low decrease in models' performances in contrast to the drop caused by the 5–10 m error. For example, in the case of MaxEnt models for intermediate species, AUC dropped on average from 0.91 (unaltered models) to 0.79 for the positional error of magnitude inherent to any occurrence data (i.e. up to 10 m), and to 0.71 in the case of the extreme positional error (100–500 m), respectively (Fig. 2). Nevertheless, the magnitude of the negative effect of positional error varied according to the species niche breadth. For both GLM and MaxEnt models the drop between unaltered models and the smallest simulated positional error (5–10 m) was higher for specialist and intermediate species (AUC dropped on average about 0.12) than for generalist species (AUC dropped on average about 0.05).

The results showed that the positional error in the occurrence data reduced the niche overlap in both the geographical and environmental space of both GLM and MaxEnt models. Niche overlap decreased gradually with the increasing positional error with an especially significant decrease in models' niche overlap at the extreme case of the positional error (100–500 m) (Fig. 3, 4). However, the effect of the positional error on the niche overlap varied depending on species' niche breadth. Decrease in the niche overlap was higher for specialist and intermediate species than for generalist species, especially in the geographical space. For example, in case of MaxEnt models, Spearman's rank correlation was reduced from 0.98 to 0.58 for the specialist and from 0.83 to 0.70 for the generalist species, respectively (Fig. 3). However, the effect of the positional error was not that evident from I, especially for the generalist species in geographical space. For example, the decrease for generalist species and MaxEnt models was on average only from 0.96 to 0.9 and the GLM models appeared as not being affected at all.

Finally, independently of the validation metric, results showed that increasing the sample size cannot compensate for the effect of positional error (Fig. 2–4). On the contrary, it is evident that a combination of low sample size of 30 samples

with positional error led to erratic behaviour and generally low performance of the models.

### Comparison of the relative importance of individual predictors ( $R^2$ )

The results show that the positional error and modelling technique had the highest relative importance ( $R^2$ ) for the model performance (AUC, TSS). The relative importance of the sample size and niche breadth was much smaller and mutually comparable (Table 2). According to the niche overlap in geographical space assessed by I (model predictions), niche breadth had the greatest effect, followed by the positional error, modelling technique and sample size, the importance of which was almost negligible. In contrast, according to correlations, the modelling technique and positional error had the highest relative importance ( $R^2$ ) followed by the niche breadth and by sample size, the importance of which was minimal. When assessing relative importance for niche overlap in the environmental space, the modelling technique and positional error showed the highest contribution followed by the niche breadth and by sample size, the importance of which was almost negligible, just like in the above metrics. All those factors significantly affected SDMs performance and predictions ( $p$ -value < 0.05).

### Discussion

In this study, we focused on the effect of positional error in species occurrences on fine-scale SDMs. We simulated species with different levels of niche breadth to assess whether there was a link between the width of the environmental niche and the effect of the size of positional error. Our results showed that introducing positional error into species occurrence data led to a decrease in model performance and prediction accuracy in both the geographical and environmental space. However, the effect of the positional error varied with species niche breadth. The same positional error had a greater impact on specialist (low ROA and prevalence, narrow breadth of niche) than on generalist (high ROA and prevalence, wide breadth of niche) species. This is likely because in case of specialist species, occurrences could be easily shifted to inappropriate environments outside of the species' environmental niche. This could also explain the inconsistent conclusions of previous studies (Graham et al. 2008, Fernandez et al. 2009).

Higher sample sizes slightly improved unaltered models' accuracy; the results however showed that increasing the sample size could not compensate for the effect of positional error on models' accuracy (Fig. 2–4). On the other hand, low sample sizes of positionally inaccurate data were especially problematic for modelling. These results are in general agreement with the study by Mitchell et al. (2017) who investigated the influence of sample size (ranging from 100 samples to 400) in conjunction with the positional error; their results showed that models based on smaller sample sizes were more affected by a positional error than those with higher numbers of species occurrences. However, it is difficult to conclude whether



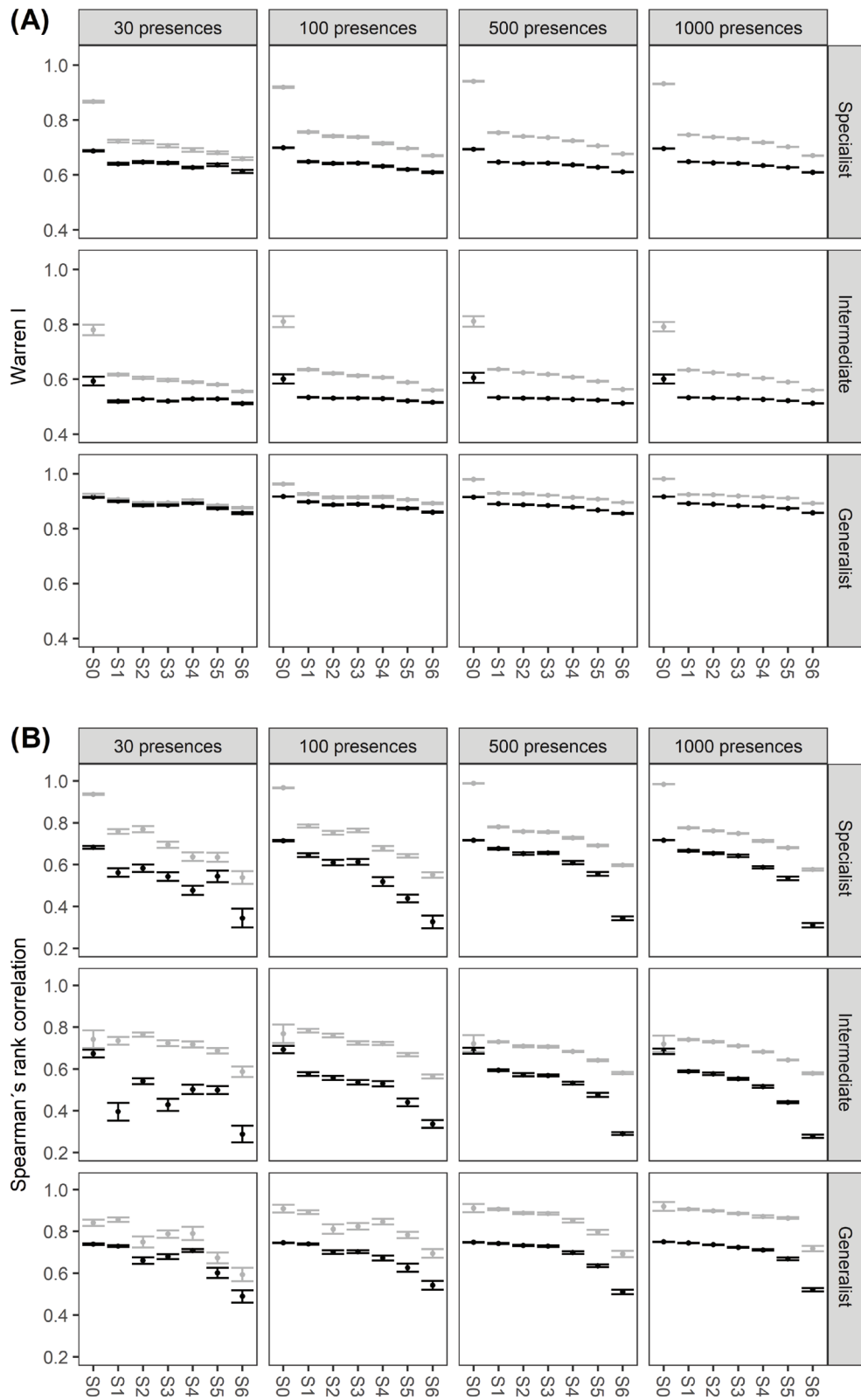


Figure 3. Resulting I (A) and Spearman's rank correlation (B) scores of niche overlap in geographical space according to different species niche breadth (specialist, intermediate, generalist), positional error (S0, unaltered models; S1, 5–10 m; S2, 10–15 m; S3, 15–20 m; S4, 20–50 m; S5, 50–100 m; S6, 100–500 m) and sample sizes (number of presences=30, 100, 500, 1000; note that for GLM models twice as many absences compared to presences were generated). Black colour shows results for GLM models while grey shows results for MaxEnt models.

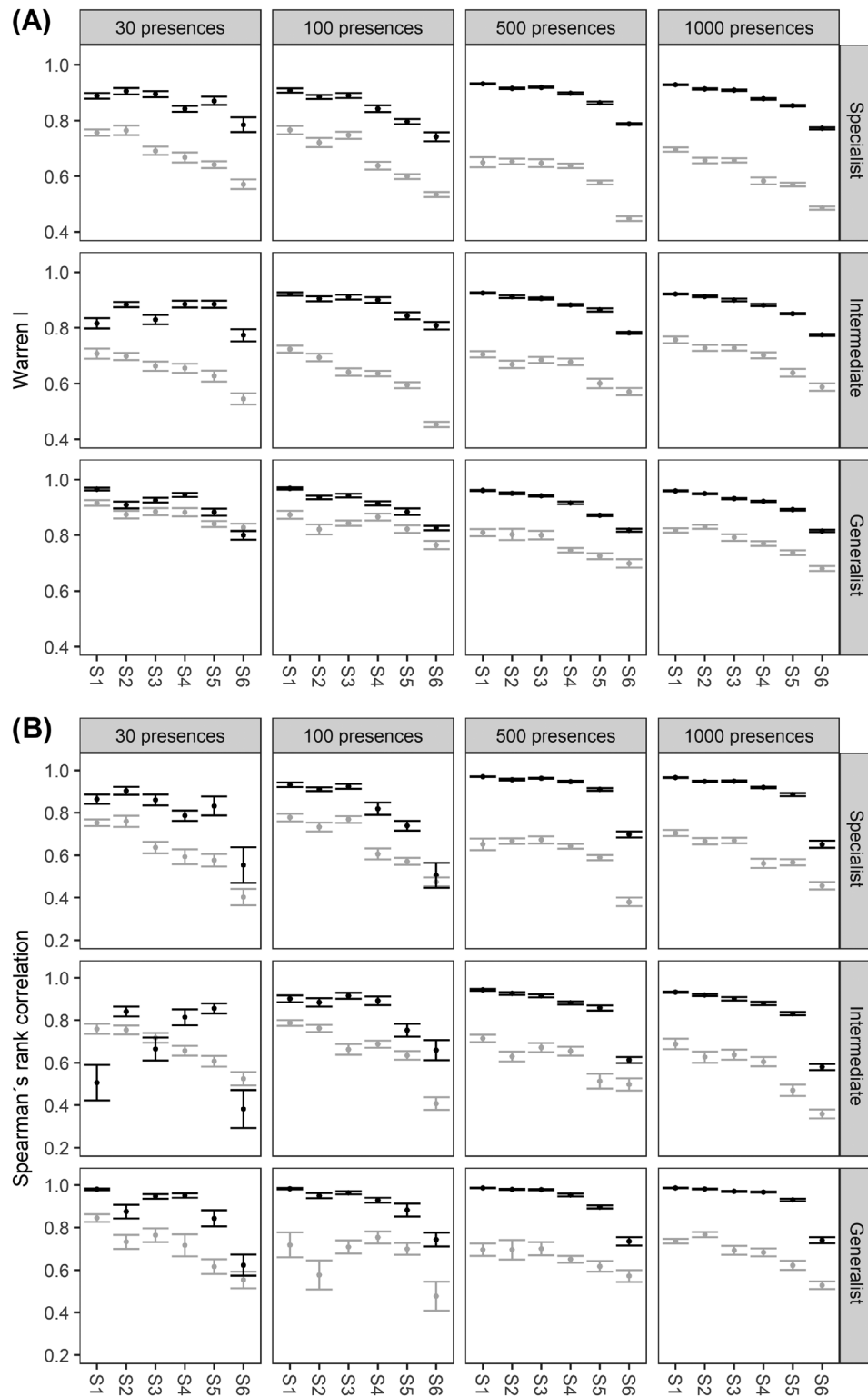


Figure 4. Resulting I (A) and Spearman's rank correlation (B) scores of niche overlap in the environmental space according to different species niche breadth (specialist, intermediate, generalist), positional error and sample size (number of presences=30, 100, 500, 1000; note that for GLM models, twice as many absences as presences were generated). Also note that here we show the niche overlap between unaltered models and models affected by a specified positional error (and not a comparison with simulated probability of occurrences as in Fig. 3). Thus, for example, S1 shows a comparison of niche overlap between unaltered models and models affected with positional error in the range of 5–10 m. Black colour shows results for GLM models while grey shows results for MaxEnt models.

Table 2. Comparison of the relative importance of individual factors ( $R^2$ , %) for ANOVA of performance metrics (AUC, TSS) and niche overlap in the geographical and environmental spaces (I, correlation).

Factor	AUC	TSS	I geographical space	Correlation geographical space	I environmental space	Correlation environmental space
ROA	4	4.14	75	11.2	9.7	1.7
Sample size	1.1	1.78	0.1	1	0.2	0.4
Modelling technique	18.7	21.35	8	24.7	45.4	21.5
Positional error	25.4	24.58	8.4	27.5	13.2	18.3

or not 100 records with positional error of 10 m are better or worse for modelling at the scale of 5 m than 500 records with positional error 25 m. For example, Moudrý and Šímová (2012) suggested that the spatial resolution of the environmental data should be coarser than the biggest positional error of the occurrence data and Naimi et al. (2011) showed that the effect of positional error is reduced by spatial autocorrelation in environmental variables. However, the trade-off between the scale and positional error has not been thoroughly studied.

The degree of decrease between unaltered and altered models (i.e. those with positional error) differed among adopted validation metrics and assuming a sufficiently large sample size, AUC and TSS provided clear evidence of decreasing model quality. The ability of evaluation metrics to identify the magnitude of error caused by positional inaccuracies was previously discussed by Osborne and Leitão (2009). Interestingly, they found that the use of AUC for the error quantification in models affected by positional error was limited as AUC did not decrease when compared to the control models. We hypothesize that this contradiction results from confounding effects of real data used in their study (i.e. they did not use virtual species). In Osborne and Leitão (2009), the modelling algorithms were allowed to choose the best combination of environmental variables from a set of twelve variables for scenarios with different levels of positional error. Indeed, they showed that positional error led to alteration of the variables selected by the modelling algorithm. The selected variables however often failed to represent the conditions pertinent to the species during habitat selection. In contrast, here we used the same variables throughout, both to generate the virtual species and to model their distribution. Hence, our modelling approaches (GLM, MaxEnt) did not have the option to select variables that would provide a closer fit to the altered occurrence data but that were lacking ecological relevance and as a result did not lead to spurious increase in AUC and TSS values. We suggest that the effect of positional error on selection of environmental variables should be further investigated.

The effects discussed above raise serious concerns as it is possible that the use of positionally inaccurate data combined with an arbitrary selection of environmental variables that may lack ecological relevance results in seemingly accurate but entirely wrong models. For instance, Fourcade et al. (2018) successfully fitted SDMs with non-ecological variables such as paintings to demonstrate this point. While Osborne and Leitão (2009) and Mitchell et al. (2017) suggested that useful predictions can still be generated from data affected by positional error, they warned that the ecological

interpretation of such data and predictions was dangerous. Our results support the importance of assessing data in terms of fitness-for-use (Lecours 2017). Fitness-for-use is the concept of determining whether or not a dataset is of sufficient quality for a particular purpose (Goodchild 2006). Spatial scale is intrinsically linked to such assessment of fitness-for-use (Lecours et al. 2017) as data accuracy is dependent on the spatial resolution of the environmental data. As indicated by Moudrý and Šímová (2012), the spatial resolution of the environmental data should always be coarser than the largest positional error associated with occurrence data.

In line with previous work (Van Niel and Austin 2007, Rocchini et al. 2011, Lecours et al. 2017), we believe that attempts to predict species distributions with data of unknown accuracy are potentially dangerous and as such, we highlight the necessity of quantifying the positional accuracy of data. If such assessment is limited by metadata availability, for example in case of historical data, we recommend to at least approximate the positional accuracy based on known information such as the collection methodology or the number of decimals recorded with coordinates. With a proper fitness-for-use assessment that includes data quality and scale, the resolution of environmental variables can be coarsened before they are integrated into a modelling exercise to minimize the adverse effects of the positional error of species occurrences. However, we are aware that this may involve altering the spatial resolution of data to a level that is no longer eligible for potentially optimal resolution(s), i.e. the scale at which species respond to the environment (Lecours et al. 2015, Moudrý et al. 2019). As demonstrated in Lecours et al. (2017), there is a trade-off between spatial scale and data quality that needs to be evaluated as a part of the fitness-for-use assessment. While no experiments are currently available to help quantify which is more important for successful modelling (whether it is the data quality or scale), we suggest that pre-analyses be performed to test whether keeping a finer resolution is more important than minimizing positional error, or vice-versa. For new surveys, we suggest paying a close attention to measurement techniques to minimize positional error, for instance by using differential GNSS, especially for species with a narrow ecological niche as our results show that the positional error of species occurrence data has a profound effect on results of SDMs. Finally, we advocate for additional studies focused on the influence of positional error using more complex virtual species (e.g. with a higher number of environmental variables or with more complex response curves) to improve SDM use in ecology, macroecology and biogeography.

## Conclusions

In this study, we explored how positional error in species occurrences affects fine-scale SDMs. We showed that the influence of positional error on SDMs differed according to the width of species' ecological niches and this effect was evident in both geographical and environmental space. The effect of the positional error on generalist species was much smaller than the effect on specialist species, which were affected the most. In addition, our results show that the negative effects of positionally inaccurate data entering SDMs cannot be mitigated by increasing the sample size. Therefore, a take away message of our study is that improving positional accuracy of data appears to be more effective than increasing sample size. We suggest that it is critical to evaluate the quality of data with respect to the spatial resolution of the environmental variables and to select occurrences with a low positional error (note that a low positional error can be even 1 km if the spatial resolution of environmental variables is of similar size). Future research should be focused on the influence of positional error using more complex virtual species (e.g. with a higher number of environmental variables or with more complex response curves) and on how positional accuracy errors may affect the selection of variables in modelling species distribution to improve its future application in ecology, macroecology and biogeography.

## Data availability statement

Using our methods, species occurrence data may be artificially generated using virtualspecies package in R. The LiDAR data are owned by Krkonose Mountains National Park and are available upon request for research purposes.

*Acknowledgements* – The authors would like to thank the Krkonose Mountains National Park for providing LiDAR data. We greatly appreciate the contribution of the subject editor and both reviewers. *Funding* – This research was funded by the Internal Grant Agency of Faculty of Environmental Sciences, Czech Univ. of Life Sciences Prague, grant no. 20174241 and no. 20194224. VM, VB and MF were also supported by the Czech Science Foundation (project no. 17-17156Y).

*Author contributions* – All authors contributed substantially to the work. VM and TV are authors of the main idea of the research and supervised the whole research. LG and VB performed all GIS and statistical analyses. VB supervised statistical analyses. MF processed LiDAR data. LG wrote the first draft of the manuscript. VL, MM, PŠ and DR helped to improve the manuscript. All authors gave final approval for publication.

## References

- Allouche, O. et al. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). – *J. Appl. Ecol.* 43: 1223–1232.
- Araújo, M. B. et al. 2019. Standards for distribution models in biodiversity assessments. – *Sci. Adv.* 5: eaat4858.
- Arribas, P. et al. 2012. Dispersal ability rather than ecological tolerance drives differences in range size between lentic and lotic water beetles (Coleoptera: Hydrophilidae). – *J. Biogeogr.* 39: 984–994.
- Barbet-Massin, M. et al. 2012. Selecting pseudo-absences for species distribution models: how, where and how many? – *Methods Ecol. Evol.* 3: 327–338.
- Besnard, A. G. et al. 2013. Topographic wetness index predicts the occurrence of bird species in floodplains. – *Divers. Distrib.* 19: 955–963.
- Beven, K. J. and Kirkby, M. J. 1979. A physically based, variable contributing area model of basin hydrology. – *Hydrol. Sci. J.* 24: 43–69.
- Boulangeat, I. et al. 2012. Niche breadth, rarity and ecological characteristics within a regional flora spanning large environmental gradients. – *J. Biogeogr.* 39: 204–214.
- Breiner, F. T. et al. 2015. Overcoming limitations of modelling rare species by using ensembles of small models. – *Methods Ecol. Evol.* 6: 1210–1218.
- Brown, J. H. 1984. On the relationship between abundance and distribution of species. – *Am. Nat.* 124: 255–279.
- Bulluck, L. et al. 2006. Spatial and temporal variations in species occurrence rate affect the accuracy of occurrence models. – *Global Ecol. Biogeogr.* 15: 27–38.
- Chefaoui, R. M. et al. 2011. Effects of species' traits and data characteristics on distribution models of threatened invertebrates. – *Anim. Biodivers. Conserv.* 34: 229–247.
- Cianfrani, C. et al. 2018. More than range exposure: global otter vulnerability to climate change. – *Biol. Conserv.* 221: 103–113.
- Connor, T. et al. 2018. Effects of grain size and niche breadth on species distribution modeling. – *Ecography* 41: 1270–1282.
- Conrad, O. 2003. Module topographic wetness index (SAGA). – Version 2.1.3.
- Coops, N. C. et al. 2010. Assessing the utility of LiDAR remote sensing technology to identify mule deer winter habitat. – *Can. J. Remote Sens.* 36: 81–88.
- Davies, A. B. and Asner, G. P. 2014. Advances in animal ecology from 3D-LiDAR ecosystem mapping. – *Trends Ecol. Evol.* 29: 681–691.
- Devictor, V. et al. 2010. Defining and measuring ecological specialization. – *J. Appl. Ecol.* 47: 15–25.
- Eaton, S. et al. 2018. Adding small species to the big picture: species distribution modelling in an age of landscape scale conservation. – *Biol. Conserv.* 217: 251–258.
- Evangelista, P. H. et al. 2008. Modelling invasion for a habitat generalist and a specialist plant species. – *Divers. Distrib.* 14: 808–817.
- Fernandes, R. F. et al. 2018. How much should one sample to accurately predict the distribution of species assemblages? A virtual community approach. – *Ecol. Inform.* 48: 125–134.
- Fernandez, M. et al. 2009. Locality uncertainty and the differential performance of four common niche-based modeling techniques. – *Biodivers. Inform.* 6: 36–52.
- Ferrier, S. et al. 2017. Biodiversity modelling as part of an observation system. – In: Walters, M. and Scholers, R. (eds), *The GEO handbook on biodiversity observation networks*. Springer, pp. 239–257.
- Fielding, A. H. and Bell, J. F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. – *Environ. Conserv.* 24: 38–49.
- Fourcade, Y. et al. 2014. Mapping species distributions with MAXENT using a geographically biased sample of presence data: a

- performance assessment of methods for correcting sampling bias. – *PLoS One* 9: e97122.
- Fourcade, Y. et al. 2018. Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. – *Global Ecol. Biogeogr.* 27: 245–256.
- Frair, J. L. et al. 2010. Resolving issues of imprecise and habitat-biased locations in ecological analyses using GPS telemetry data. – *Phil. Trans. R. Soc. B* 365: 2187–2200.
- Franklin, J. 2010. Mapping species distributions: spatial inference and prediction. – Cambridge Univ. Press.
- Franklin, J. et al. 2014. Linking spatially explicit species distribution and population models to plan for the persistence of plant species under global change. – *Environ. Conserv.* 41: 97–109.
- Futuyma, D. J. and Moreno, G. 1988. The evolution of ecological specialisation. – *Annu. Rev. Ecol. Syst.* 207–233.
- Gábor, L. et al. 2019. How do species and data characteristics affect species distribution models and when to use environmental filtering? – *Int. J. Geogr. Inform. Sci.* doi: 10.1080/13658816.2019.1615070
- Gaston, K. J. et al. 1997. Interspecific abundance range size relationships: an appraisal of mechanisms. – *J. Anim. Ecol.* 66: 579–601.
- Goodchild, M. F. 2006. Fundamentals of spatial data quality. – ISTE, London.
- Gottschalk, T. K. et al. 2011. Influence of grain size on species–habitat models. – *Ecol. Model.* 222: 3403–3412.
- Graham, C. H. et al. 2008. The influence of spatial errors in species occurrence data used in distribution models. – *J. Appl. Ecol.* 45: 239–247.
- Hijmans, R. J. 2012. Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. – *Ecology* 93: 679–688.
- Holloway, P. et al. 2016. Incorporating movement in species distribution models: how do simulations of dispersal affect the accuracy and uncertainty of projections? – *Int. J. Geogr. Inform. Sci.* 30: 2050–2074.
- Jiménez-Valverde, A. 2012. Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. – *Global Ecol. Biogeogr.* 21: 498–507.
- Johnson, C. J. and Gillingham, M. P. 2008. Sensitivity of species-distribution models to error, bias and model design: an application to resource selection functions for woodland caribou. – *Ecol. Model.* 213: 143–155.
- Khosravipour, A. et al. 2016. Generating spike-free digital surface models using LiDAR raw point clouds: a new approach for forestry applications. – *Int. J. Appl. Earth Observ. Geoinform.* 52: 104–114.
- Kopecký, M. and Čížková, Š. 2010. Using topographic wetness index in vegetation ecology: does the algorithm matter? – *Appl. Veg. Sci.* 13: 450–459.
- Lecours, V. 2017. On the use of maps and models in conservation and resource management (warning: results may vary). – *Front. Mar. Sci.* 4: 1–18.
- Lecours, V. et al. 2015. Spatial scale and geographic context in benthic habitat mapping: review and future directions. – *Mar. Ecol. Progr. Ser.* 535: 259–284.
- Lecours, V. et al. 2017. Artefacts in marine digital terrain models: a multiscale analysis of their impact on the derivation of terrain attributes. – *IEEE Trans. Geosci. Remote Sens.* 55: 5391–5406.
- Lefsky, M. A. et al. 2002. LiDAR remote sensing for ecosystem studies: LiDAR, an emerging remote sensing technology that directly measures the three-dimensional distribution of plant canopies, can accurately estimate vegetation structural attributes and should be of particular interest to forest, landscape and global ecologists. – *BioScience* 52: 19–30.
- Leroy, B. et al. 2016. virtualspecies, an R package to generate virtual species distributions. – *Ecography* 39: 599–607.
- Leroy, B. et al. 2018. Without quality presence–absence data, discrimination metrics such as TSS can be misleading measures of model performance. – *J. Biogeogr.* 45: 1994–2002.
- Linda, R. et al. 2016. Developing a criterion for distinguishing tetraploid birch species from diploid and modelling their potential distribution on the Czech Republic. – In: Kacálek, D. et al. (eds), *Proceedings of central European silviculture*, pp. 71–77.
- Lobo, J. M. 2008. More complex distribution models or more representative data? – *Biodivers. Inform.* 5: 14–19.
- Luoto, M. et al. 2005. Uncertainty of bioclimate envelope models based on the geographical distribution of species. – *Global Ecol. Biogeogr.* 14: 575–584.
- MacKinnon, J. G. and White, H. 1985. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. – *J. Economet.* 29: 305–325.
- Malvasi, M. et al. 2018. Plant invasions in Italy: an integrative approach using the European LifeWatch infrastructure database. – *Ecol. Indic.* 91: 182–188.
- McPherson, J. M. and Jetz, W. 2007. Effects of species' ecology on the accuracy of distribution models. – *Ecography* 30: 135–151.
- Meynard, C. N. and Kaplan, D. M. 2012. The effect of a gradual response to the environment on species distribution modeling performance. – *Ecography* 35: 499–509.
- Meynard, C. N. and Kaplan, D. M. 2013. Using virtual species to study species distributions and model performance. – *J. Biogeogr.* 40: 1–8.
- Meynard, C. N. et al. 2019. Testing methods in species distribution modelling using virtual species: what have we learnt and what are we missing? – *Ecography* doi: 10.1111/ecog.04385
- Mitchell, P. J. et al. 2017. Sensitivity of fine-scale species distribution models to locational uncertainty in occurrence data across multiple sample sizes. – *Methods Ecol. Evol.* 8: 12–21.
- Moudrý, V. 2015. Modelling species distributions with simulated virtual species. – *J. Biogeogr.* 42: 1365–1366.
- Moudrý, V. and Šímová, P. 2012. Influence of positional accuracy, sample size and scale on modelling species distributions: a review. – *Int. J. Geogr. Inform. Sci.* 26: 2083–2095.
- Moudrý, V. and Šímová, P. 2013. Relative importance of climate, topography and habitats for breeding wetland birds with different latitudinal distributions in the Czech Republic. – *Appl. Geogr.* 44: 165–171.
- Moudrý, V. et al. 2017. Which breeding bird categories should we use in models of species distribution? – *Ecol. Indic.* 74: 526–529.
- Moudrý, V. et al. 2018. On the use of global DEMs in ecological modelling and the accuracy of new bare-earth DEMs. – *Ecol. Model.* 383: 3–9.
- Moudrý, V. et al. 2019. Potential pitfalls in rescaling digital terrain model-derived attributes for ecological studies. – *Ecol. Inform.* 54: 100987.
- Naimi, B. et al. 2011. Spatial autocorrelation in predictors reduces the impact of positional uncertainty in occurrence data on species distribution modelling. – *J. Biogeogr.* 38: 1497–1509.
- Nelder, J. A. and Baker, R. J. 1972. Generalized linear models. – Wiley.

- Oksanen, J. and Minchin, P. R. 2002. Continuum theory revisited: what shape are species responses along ecological gradients? – *Ecol. Model.* 157: 119–129.
- Osborne, P. E. and Leitão, P. J. 2009. Effects of species and habitat positional errors on the performance and interpretation of species distribution models. – *Divers. Distrib.* 15: 671–681.
- Peers, M. J. et al. 2012. Reconsidering the specialist–generalist paradigm in niche breadth dynamics: resource gradient selection by Canada lynx and bobcat. – *PLoS One* 7: e51488.
- Phillips, S. J. et al. 2006. Maximum entropy modelling of species geographic distributions. – *Ecol. Model.* 190: 231–259.
- Phillips, S. J. et al. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. – *Ecol. Appl.* 19: 181–197.
- Qiao, H. et al. 2015. Marble algorithm: a solution to estimating ecological niches from presence-only records. – *Sci. Rep.* 5: 14232.
- Quinn, P. F. B. J. et al. 1991. The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. – *Hydrol. Process.* 5: 59–79.
- Ranc, N. et al. 2016. Performance tradeoffs in target-group bias correction for species distribution models. – *Ecography* 40: 1076–1087.
- Ratray, A. et al. 2014. Quantification of spatial and thematic uncertainty in the application of underwater video for benthic habitat mapping. – *Mar. Geodesy* 37: 315–336.
- Reif, J. et al. 2018. Competition-driven niche segregation on a landscape scale: evidence for escaping from syntopy towards allotopy in two coexisting sibling passerine species. – *J. Anim. Ecol.* 87: 774–789.
- Rocchini, D. et al. 2011. Accounting for uncertainty when mapping species distributions: the need for maps of ignorance. – *Progr. Phys. Geogr.* 35: 211–226.
- Rödger, D. and Engler, J. O. 2011. Quantitative metrics of overlaps in Grinnellian niches: 422 advances and possible drawbacks. – *Global Ecol. Biogeogr.* 20: 915–927.
- Šimová, P. et al. 2019. Fine scale waterbody data improve prediction of waterbird occurrence despite coarse species data. – *Ecography* 42: 511–520.
- Slatyer, R. A. et al. 2013. Niche breadth predicts geographical range size: a general ecological pattern. – *Ecol. Lett.* 16: 1104–1114.
- Tingley, R. et al. 2018. Integrating transport pressure data and species distribution models to estimate invasion risk for alien stowaways. – *Ecography* 41: 635–646.
- Václavík, T. and Meentemeyer, R. K. 2012. Equilibrium or not? Modelling potential distribution of invasive species in different stages of invasion. – *Divers. Distrib.* 18: 73–83.
- Van Niel, K. P. and Austin, M. P. 2007. Predictive vegetation modelling for conservation: impact of error propagation from digital elevation data. – *Ecol. Appl.* 17: 266–280.
- Vogeler, J. C. et al. 2014. Terrain and vegetation structural influences on local avian species richness in two mixed-conifer forests. – *Remote Sens. Environ.* 147: 13–22.
- Warren, D. L. et al. 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. – *Evolution* 62: 2868–2883.
- Warren, D. L. et al. 2019a. Evaluating species distribution models with discrimination accuracy is uninformative for many applications. – *BioRxiv* 684399.
- Warren, D. L. et al. 2019b. danlwarren/ENMTools: initial beta release. – Package ver. 0.2, Zenodo, <<https://github.com/danlwarren/ENMTools>>.
- Watts, S. M. et al. 2019. Modelling potential habitat for snow leopards (*Panthera uncia*) in Ladakh, India. – *PLoS One* 14: e0211509.
- Wieczorek, J. et al. 2004. The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. – *Int. J. Geogr. Inform. Sci.* 18: 745–767.
- Wisz, M. S. et al. 2008. Effects of sample size on the performance of species distribution models. – *Divers. Distrib.* 14: 763–773.
- Ye, X. et al. 2018. Impacts of future climate and land cover changes on threatened mammals in the semi-arid Chinese Altai Mountains. – *Sci. Total Environ.* 612: 775–787.
- Zeileis, A. 2006. Object-oriented computation of sandwich estimators. – *J. Stat. Softw.* 16: 1–16.
- Zhang, G. et al. 2018. A heuristic-based approach to mitigating positional errors in patrol data for species distribution modeling. – *Trans. GIS* 22: 202–216.
- Zurell, D. et al. 2010. The virtual ecologist approach: simulating data and observers. – *Oikos* 119: 622–635.

Supplementary material (available online as Appendix ecog-04687 at <[www.ecography.org/appendix/ecog-04687](http://www.ecography.org/appendix/ecog-04687)>). Appendix 1–5.