Received: 8 November 2021 | Accepted: 27 June 2022
DOI: 10.1111/2041-210X.13956

Methods in Ecology and Evolution | BRITISH ECOLOGICAL SOCIETY

RESEARCH ARTICLE

Positional errors in species distribution modelling are not overcome by the coarser grains of analysis

Lukáš Gábor[1,2,3] | Walter Jetz[2,3] | Muyang Lu[2,3] | Duccio Rocchini[1,4] | Anna Cord[5] | Marco Malavasi[1] | Alejandra Zarzo-Arias[1,6,7] | Vojtěch Barták[1] | Vítězslav Moudrý[1]

[1]Department of Spatial Sciences, Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Praha – Suchdol, Czech Republic; [2]Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut, USA; [3]Center for Biodiversity and Global Change, Yale University, New Haven, Connecticut, USA; [4]BIOME Lab, Department of Biological, Geological and Environmental Sciences, Alma Mater Studiorum University of Bologna, Bologna, Italy; [5]Institute of Geography, Technische Universität Dresden, Dresden, Germany; [6]Universidad de Oviedo, Oviedo, Asturias, Spain and [7]Department of Biogeography and Global Change, Museo Nacional de Ciencias Naturales (MNCN-CSIC), Madrid, Spain

Correspondence
Lukáš Gábor
Email: gabor@fzp.czu.cz

Funding information
Internal Grant Agency of Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Grant/Award Number: 2020B0009; Technological grant agency of the Czech Republic, Grant/Award Number: SS02030018

Handling Editor: Phil Bouchet

Abstract

1. The performance of species distribution models (SDMs) is known to be affected by analysis grain and positional error of species occurrences. Coarsening of the analysis grain has been suggested to compensate for positional errors. Nevertheless, this way of dealing with positional errors has never been thoroughly tested. With increasing use of fine-scale environmental data in SDMs, it is important to test this assumption. Models using fine-scale environmental data are more likely to be negatively affected by positional error as the inaccurate occurrences might easier end up in unsuitable environment. This can result in inappropriate conservation actions.
2. Here, we examined the trade-offs between positional error and analysis grain and provide recommendations for best practice. We generated narrow niche virtual species using environmental variables derived from LiDAR point clouds at 5 × 5 m fine-scale. We simulated the positional error in the range of 5 m to 99 m and evaluated the effects of several spatial grains in the range of 5 m to 500 m. In total, we assessed 49 combinations of positional accuracy and analysis grain. We used three modelling techniques (MaxEnt, BRT and GLM) and evaluated their discrimination ability, niche overlap with virtual species and change in realized niche.
3. We found that model performance decreased with increasing positional error in species occurrences and coarsening of the analysis grain. Most importantly, we showed that coarsening the analysis grain to compensate for positional error did not improve model performance. Our results reject coarsening of the analysis grain as a solution to address the negative effects of positional error on model performance.

4. We recommend fitting models with the finest possible analysis grain and as close to the response grain as possible even when available species occurrences suffer from positional errors. If there are significant positional errors in species occurrences, users are unlikely to benefit from making additional efforts to obtain higher resolution environmental data unless they also minimize the positional errors of species occurrences. Our findings are also applicable to coarse analysis grain, especially for fragmented habitats, and for species with narrow niche breadth.

KEYWORDS
georeferencing, grain size, resolution, scale, SDM, virtual species

1 | INTRODUCTION

Species distribution models (SDMs) use species occurrence data and environmental explanatory variables to infer species–environment relationships and predict species distribution ranges (Ferrier et al., 2017). Despite their routine use and relatively well-established practices (Simoes et al., 2020) and standards (Araújo et al., 2019; Merow et al., 2019), some methodological considerations still require further investigation. With the increasing availability of heterogeneous data from a multitude of sources of varying quality, careful assessment of uncertainties and purpose-built methodologies are becoming more important (Wüest et al., 2020). Indeed, recent recommendations and methodological improvements are particularly relevant to data quality issues such as positional error, sampling bias, sample size and scale. Specialized tools have been developed for the identification of positionally inaccurate records (e.g. Robertson et al., 2016; Zizka et al., 2019). Similarly, development and testing of sampling bias correction methods continue (Gábor, Moudrý, Barták, & et al., 2020; Inman et al., 2021) as well as the research into the effects of sample size (Hallman & Robinson, 2020; Jiménez-Valverde, 2020; McPherson et al., 2004; McPherson & Jetz, 2007) and of changing the grain of response and explanatory variables (Mertes & Jetz, 2018; Šímová et al., 2019).

Additionally, a key question, namely at which spatial scales (grains) the ecological processes underlying species distribution patterns operate, continues to be debated (Mertes & Jetz, 2018; Miguet et al., 2016; Pearson & Dawson, 2003). SDMs can be developed on a very wide range of grains (e.g. from 1 m^2 to 10,000 km^2 or more) and several studies (e.g. Guisan et al., 2007; Kaliontzopoulou et al., 2008; Seo et al., 2009) reported effects of the analysis grain on the performance of SDMs. At some spatial scales, species respond more strongly to their environment than at others (Holland et al., 2004; Mayor et al., 2009; McGarigal et al., 2016). This is often referred to as ecological scale, scale of effect, response grain or response scale (Holland et al., 2004; Mertes & Jetz, 2018; Wu & Li, 2006). Here, we follow Mertes and Jetz (2018) and use the term 'response grain' to indicate the theoretical scale at which individuals of a species respond to environmental factors and 'analysis grain' to describe the spatial unit (grain) at which the species occurrence is modelled. As the chosen analysis grain affects our ability to detect the species' response to environmental factors (variables), factors such as positional errors of species occurrences, resolution of available environmental data and the response grain on which species are expected to respond to the environment need to be considered (Dungan et al., 2002; Lechner et al., 2012; Lecours et al., 2015; Schneider, 2001).

It is increasingly recognized that positional uncertainty (associated with the location of species observations) is an important factor to consider during the modelling process. Positional errors cause problems in modelling, as environmental conditions at the recorded locations might differ from those at actual locations, which (as was demonstrated) can have a significant impact on SDM results. For example, Visscher (2006) showed that positional error can bias inferences about species–environment relationships. Similarly, Johnson and Gillingham (2008) concluded that positional errors have a significant effect on model quality, and Osborne and Leitão (2009) recommended minimizing positional errors through careful study design and data processing. More recently, Hefley et al. (2014) pointed out that positional errors can lead to biased estimates of regression coefficient. Indeed, the Darwin Core Standard (https://dwc.tdwg.org/) has proven to be useful for recording positional uncertainty of species occurrences (Wieczorek et al., 2012), and the importance of georeferencing accuracy has been highlighted by many studies (e.g. Moudrý & Devillers, 2020), including a report on the suitability of Global Biodiversity Information Facility (GBIF) data for use in SDMs (Anderson et al., 2016).

Notably, with the increasing use of fine-scale resolution data in SDM, such as variables derived from LiDAR with a resolution of a few meters (e.g. Lecours et al., 2020; Moudrý et al., 2021; Pradervand et al., 2014; Sillero & Goncalves-Seco, 2014; Simonson et al., 2014; Wüest et al., 2020), the negative effects of positional error in species occurrence data are no longer associated only with relatively old datasets (e.g. from herbarium or museum collections), but it is also necessary to consider positional errors inherent to data georeferenced using global navigation satellite systems. Indeed, Gábor, Moudrý, Lecours, et al. (2020) used a 5 × 5 m analysis grain and reported that the largest drop in model performance was observed at the smallest simulated positional error of 5–10 m (they simulated errors up to 500 m).

Both positional error and adopted analysis grain have been intensively studied; however, despite their interconnectedness, their interactions and trade-offs are rarely systematically addressed (but see Engler et al., 2004, Montgomery et al., 2011, Cheng et al., 2021). Particularly, the trade-off between the adopted analysis grain and positional error of species occurrence data is poorly acknowledged. Typically, studies try to balance these interconnected issues based on available data and metadata (i.e. users might know the positional error of occurrences but do not know the optimal grain and vice versa). For example, researchers aim to georeference species occurrences with respect to adopted analysis grain (Ballesteros-Mejia et al., 2017) or, when using already georeferenced data, they remove imprecise occurrences (e.g. records with latitude and longitude precision lower than three decimal places or with known high positional uncertainty; Gueta & Carmel, 2016, Watcharamongkol et al., 2018, Ellis-Soto et al., 2021). Alternatively, coarsening the analysis grain can be used for correcting georeferencing errors (Engler et al., 2004; Keil et al., 2014; Moudrý & Šímová, 2012; Sillero & Barbosa, 2021; Vollering et al., 2016). These techniques, however, have a drawback: removing positionally inaccurate records or coarsening the analysis grain reduce the sample size. Moreover, the latter approach can lead to the loss of explanatory power of the model (as the grain at which species respond to the environment might be better represented by a finer grain). This may indeed limit our ability to observe how species respond to the environment (Mertes & Jetz, 2018).

All in all, it is evident that both analysis grain and positional accuracy are important and interacting factors affecting SDM results (i.e. environmental niches and spatial distributions of modelled species). However, the knowledge of how they interact and the implications for modelling practice is lacking. It is crucial to have this knowledge, especially with increasing availability of fine-scale environmental data (e.g. Haesen et al., 2021; Li et al., 2021) and their use in predictive models developed for conservation and climate change studies (see for example Lembrechts, Lenoir, et al., 2019; Lembrechts, Nijs, & Lenoir, 2019; Stark & Fridley, 2022; Zellweger et al., 2019). Therefore, we here address the following questions: (a) What are the trade-offs between analysis grain and positional error when modelling species distributions? (b) Is it advisable to coarsen the analysis grain to minimize the effect of the positional error, or should the analysis grain be kept as close as possible to the assumed response grain, regardless of the positional error?

2 | MATERIALS AND METHODS

2.1 | LiDAR data and derived environmental variables

We used a point cloud from airborne laser scanning of Krkonose Mountains National Park, Czech Republic, that covers over 370 square kilometres (approximately 30 km in west/east direction and 13 km in south/north direction), to derive three fine-scale environmental variables. It has been shown that the negative effect of positional error varies according to the degree of spatial autocorrelation in environmental variables. The lower is the spatial autocorrelation in environmental variables the more pronounced is the negative effect of positional error in species occurrences (Naimi et al., 2011, 2014). Therefore, we chose environmental variables with various levels of spatial autocorrelation to mimic a real modelling situation (Figure A1). Note, that spatial autocorrelation is a function of resolution and may change as the analysis grain is coarsened (see Mertes & Jetz, 2018). However, this is not our case, as the environmental variables maintained similar spatial autocorrelation across all used response grains (see Figure A1). Specifically, we used the canopy height model (CHM) representing structural variability of the canopy, topographic wetness index (TWI) as a surrogate for soil moisture, thus affecting vegetation composition, and altitude in the form of a digital terrain model (DTM) as a surrogate for microclimatic conditions. All these variables have been used in other studies for modelling species distributions, for example, of birds (e.g. Bakx et al., 2019; Reif et al., 2018; Vogeler et al., 2014). Hence, our virtual species might represent a bird with certain habitat requirements in terms of vegetation structure, climate and terrain characteristics. To derive the three environmental variables at a resolution of 5 × 5 m, first the point cloud was classified into vegetation, building and ground classes in the ENVI and LAStools software (Klápště et al., 2020). Second, following Khosravipour et al. (2016), we used points classified as vegetation to produce the CHM; points representing ground were used to create the DTM, which was subsequently used to derive the TWI.

2.2 | Generating virtual species

We adopted the virtual species approach, which is increasingly used to answer methodological questions related to SDMs (Zurell et al., 2010). This popularity is due in particular to the fact that it is difficult to draw clear methodological conclusions with real data, since the actual distribution as well as data deficiencies that might influence the results are unknown (Grimmett et al., 2021; Inman et al., 2021; Meynard et al., 2019; Moudrý, 2015). We used the VIRTUALSPECIES package (ver. 1.5.1) in the statistical software R (R Core Team, 2021) to generate virtual species (Leroy et al., 2016). To begin, we defined the response of virtual species to the environmental gradient at a resolution of 5 × 5 m (i.e. the finest resolution at which environmental variables were available). We used a normal distribution with the following parameters: (a) mean canopy height of 9 m and standard deviation of 4 m, (b) mean altitude of 846 m and standard deviation of 100 m and (c) mean TWI of 8 and standard deviation of 0.4 m. These parameters allowed us to simulate virtual species with a narrow niche breadth as it has been suggested that SDMs of such species are more prone to positional error (Gábor, Moudrý, Lecours, et al., 2020; Visscher, 2006). We then multiplied the responses to obtain an environmental suitability raster. We applied the probabilistic approach (logistic function with $\alpha = -0.05$ and $\beta = 0.3$) to convert the environmental suitability raster into probabilities of occurrences that were subsequently used to sample binary presence–absence rasters. We developed both presence-only and presence–absence models (see below), using 99 presence sites and 200 absence sites (i.e. sample prevalence of 0.33), and a uniform random distribution for sampling species presences and absences. The virtual species could be recreated using the 'vs' object and R script that is available via the Dryad repository (see the data availability statement for link).

2.3 | Simulating positional error and coarsening the analysis grain

Positional error in species occurrence data may range from a few metres up to hundreds of metres, depending on the data gathering technique and the source of the error. Here, we simulated the positional error in the range of 5 m to 99 m. We shifted each occurrence point in a random direction by a specified distance according to 6 scenarios. Each scenario is associated with a different shift, as follows: S1: 5–9 m, S2: 10–19 m, S3: 20–29 m, S4: 30–39 m, S5: 40–49 m and S6: 90–99 m. The scenario with the original, that is, not shifted, data is referred to as 'unaltered' hereafter. The R functions we used to simulate positional error in species occurrences are available in the R script via the Dryad repository. To test the effect of coarsening the analysis grain and, in particular, to assess whether the coarsening of the analysis grain can compensate for the negative effect of the positional error, we ran models at seven analysis grains representing two distinct situations, namely: (a) the response grain is known and relatively fine-scale data are available (5 ×5 m, 20×20 m, 40×40 m, 60×60 m, 80×80 m and 100×100 m) and (b) the analysis grain is selected on the basis of data availability (500×500 m). In the first situation, we used small steps (changes) and multiple scales to capture any minor changes, whereas in the second situation, the analysis was conducted with a grain considerably coarser than the response grain (a hundred times coarsened grain), which is undoubtedly a situation prevalent in current modelling practice. Thus, a total of 49 combinations of positional accuracy of species occurrences and analysis grains were evaluated. All environmental variables were resampled to coarser grains using the mean values of the original data (Moudrý et al., 2019). Note that coarsening the analysis grain results in multiple sampling sites ending up in the same cell (e.g. Engler et al., 2004; Guisan et al., 2007). When absences and presences occurred in the coarser grain cell after aggregation, the cell was considered a 'presence' cell, resulting in a small decrease in the number of absences. We did not observe multiple presences aggregated into a single cell (note that the largest analysis grain also limited the maximum number of background points for MaxEnt; see Table A1). It is intuitive that the quality of the models is related to sample size. Indeed, prior studies showed that sample size play an important role in SDMs. In particular, they mostly concentrated on the effects of available presences on the development of accurate presence-only models (e.g. van Proosdij et al., 2016; Wisz et al., 2008). Recently, Liu et al. (2019) used virtual species approach and recommended that hundreds of presences are needed to reach the plateau where increasing the sample size adds little to the model performance. Therefore, we keep constant number of 99 presences for all scenarios. McPherson et al. (2004) evaluated the effects of sample size on the development of presence–absence models and shown that models trained with sample size of 300 (presences and absences) perform better than those trained with 100. In addition, Jiménez-Valverde et al. (2009) found that the effect of the sample size becomes apparent for models trained with less than 70 samples. Therefore, for presence–absence models we keep the constant number of 99 presences, and we let the absences to slightly vary between 150 and 200 (Table A1). Such minimal changes in number of absences certainly did not affect our results.

2.4 | Model fitting

Three common modelling methods were used to fit species occurrence to environmental predictors: generalized linear model (GLM), boosted regression tree (BRT) and the maximum entropy model (MaxEnt). GLM, implemented in the R package GLM2 (ver. 1.2.1, Nelder & Wedderburn, 1972; Oksanen & Minchin, 2002), and BRTs, implemented in the GBM package (ver. 2.1.5, Friedman et al., 2000), represented presence–absence methods, and MaxEnt, implemented in the DISMO package (ver. 1.1-4, Phillips et al., 2006; ver. 3.4.3 of maxent.jar file, Phillips et al., 2020), a presence-background method. Using both presence-absence and presence-background methods allowed us to assess whether they are equally affected by positional errors and by coarsening of the analysis grain. The GLM was run with a logit link function and a binomial distribution. The quadratic terms of the environmental variables were included based on the known normal distribution curves of the response function. For BRT, we used Bernoulli distribution, shrinkage (learning rate) of 0.01, tree complexity of 1 (i.e. without interaction terms), bag fraction (the proportion of data used when selecting optimal tree number) of 0.5, and the maximum number of trees of 5,000. MaxEnt was used with default settings (i.e. auto features, logistic output format) and 10,000 backgrounds points. The only exception was for models with an analysis grain of 500×500 m, where the number of grids/cells was not sufficient to sample 10,000 background points, so we ended up with a smaller number of background points (see Table A1). The same three environmental variables (CHM, DTM and TWI) that were used in the process of generating virtual species were also used to fit the models in seven analysis grains (see the previous section).

2.5 | Model evaluation

We used several discrimination metrics to evaluate the performance of the models. First, we used the Sørensen index (SI), which has been recommended for the evaluation of experiments testing SDM methodologies using virtual species (Leroy et al., 2018; Li & Guo, 2013). We also aimed to determine whether predictions using erroneous/altered data tend to over- or underpredict species occurrences. Thus, we calculated the overprediction and underprediction rates. Overprediction refers to the proportion of observed absences in the predicted presence area, and underprediction measures the proportion of actual presences that were not predicted by the model (Barbosa et al., 2013; Leroy et al., 2018). However, these metrics use only three components (true positives, false positives and false negatives) of the confusion matrix and neglect the prediction of true negatives (Leroy et al., 2018). Because we manipulated the input data (i.e. introduced the positional error and changed the analysis grain), we were concerned that this might also affect the true negatives. Therefore, we added the area under the receiver operating characteristic curve (AUC; Fielding & Bell, 1997; despite recent criticisms of this metric, see for example Lobo et al., 2008, Jiménez-Valverde, 2012) and the true skill statistics (TSS; Allouche et al., 2006), which are commonly used to assess the discriminatory power of models.

In addition, we took advantage of the virtual species approach and compared differences between the predicted distribution inferred from the models and the true probability of occurrence of virtual species in geographical space. However, it has been stressed that metrics used for niche comparison are seriously affected by the inclusion of large number of cells where the species are absent (i.e. with low occurrence probabilities), and it has been recommended to remove such cell from the evaluation (Rödder & Engler, 2011). Therefore, for this evaluation, we extract occurrence probability only for occurrence data, which were used in the models. We used Spearman's rank correlation to quantify the differences. See Figure A2 for visual comparison between virtual species true distribution and predicted probability of all modelled scenarios. Note that this comparison was performed using the same resolution for all models' predictions (i.e. 500 m).

The model performance was evaluated at the analysis grain at which the individual models were fitted, which is a common practise in studies evaluating effect of analysis grain on the performance of SDM (e.g. Guisan et al., 2007, Kaliontzopoulou et al., 2008, Seo et al., 2009, Mertes & Jetz, 2018; Lembrechts, Lenoir, et al., 2019; Stark & Fridley, 2022; Zellweger et al., 2019). Performance metrics for each model were calculated using fivefold cross-validation for which the data were randomly divided into fifths. Four-fifths of the data were used to train the model and the remaining one-fifth was used to assess the performance. We performed the entire process from species generation to model evaluation 50 times and calculated average values and confidence intervals (MacKinnon & White, 1985) of validation metrics from all replications. See Figure 1 for an overview of the general modelling process. Besides comparison of models' performance, we used linear regression to quantify how introducing positional error and coarsening of environmental variables affects species realized niche.

3 | RESULTS

3.1 | Effects of positional error and analysis grain on species realized niche

Figure 2 shows linear regression line plots of species realized niche for unaltered and altered occurrence data across various analysis grains and all combinations of environmental data. It is obvious, that both introducing positional error and coarsening the analysis grain led to changes in species realized niche. More notably, the coarsening of analysis grain did not help to reconstruct the original niche. The change in realized niche is more pronounced for combination of environmental variables with lower spatial autocorrelation (i.e. TWI versus CHM; see Figure A1).

3.2 | Overall model performance

All metrics largely followed the same pattern. Therefore, we focus only on SI and Spearman's rank correlation (for AUC TSS, overprediction rate and underprediction rate values, see Supporting Information Figures A3 and A4). BRT and MaxEnt performed very well while GLM performed slightly worse using unaltered data and resolution of environmental variables (5 × 5 m). The SIs of the unaltered models were 0.76 for MaxEnt, 0.74 for BRT and 0.67 for GLM (Figure 3). Spearman's rank correlation indicates that MaxEnt and BRT models using unaltered data have high niche overlap with virtual species. They reached Spearman's rank correlation of 0.95 and 0.9, respectively. In contrast GLM achieved lower niche overlap and Spearman's rank correlation of 0.6 (Figure 3).

3.3 | Effects of positional error and analysis grain

The performance of all modelling methods was negatively affected by the positional error in species occurrences. Results show a clear trend of decreasing model performance and increasing overprediction and underprediction rate with increasing positional error (Figure 3, A3), with the largest drop in performance occurring once positional error was introduced (i.e. between the no-error and 5–9 m error categories). For example, SI dropped from 0.76 to 0.72 and from 0.74 to 0.67 for MaxEnt and BRT, respectively (Figure 3). As the position error continued to increase, a slow but gradual decline in model performance was observed. The exception from this pattern is GLM modelling method where the negative effect of positional error is noticeable only for scenarios with more pronounced positional error (i.e. 40 m and higher). The SI dropped from 0.67 (unaltered models) to 0.64 (90–99 m error). Regardless of modelling technique introducing positional error led to decrease in niche overlap between true and predicted species distribution probability. For example, Spearman's rank correlation dropped from 0.96 to 0.76 for MaxEnt, respectively, from to 0.6 to 0.34 for GLM (Figure 3).

The results also show a clear trend of decreasing model performance as the analysis grain is coarsened compared with the response grain (i.e. from the original resolution at which the virtual species were generated; 5 × 5 m). The largest decrease was observed between the unaltered models (5 m) and the models with the smallest change in the analysis grain (20 m). For example, SI decreased from 0.76 to 0.72 and from 0.74 to 0.67 for MaxEnt and BRT, respectively (Figure 3). Further coarsening of the analysis grain resulted in an additional decrease in models' performance; however, the overall decrease in performance between 20 m and 500 m was less than the decrease caused by the initial change in analysis grain (Figure 3). The same pattern shows also niche comparison assessed by Spearman's

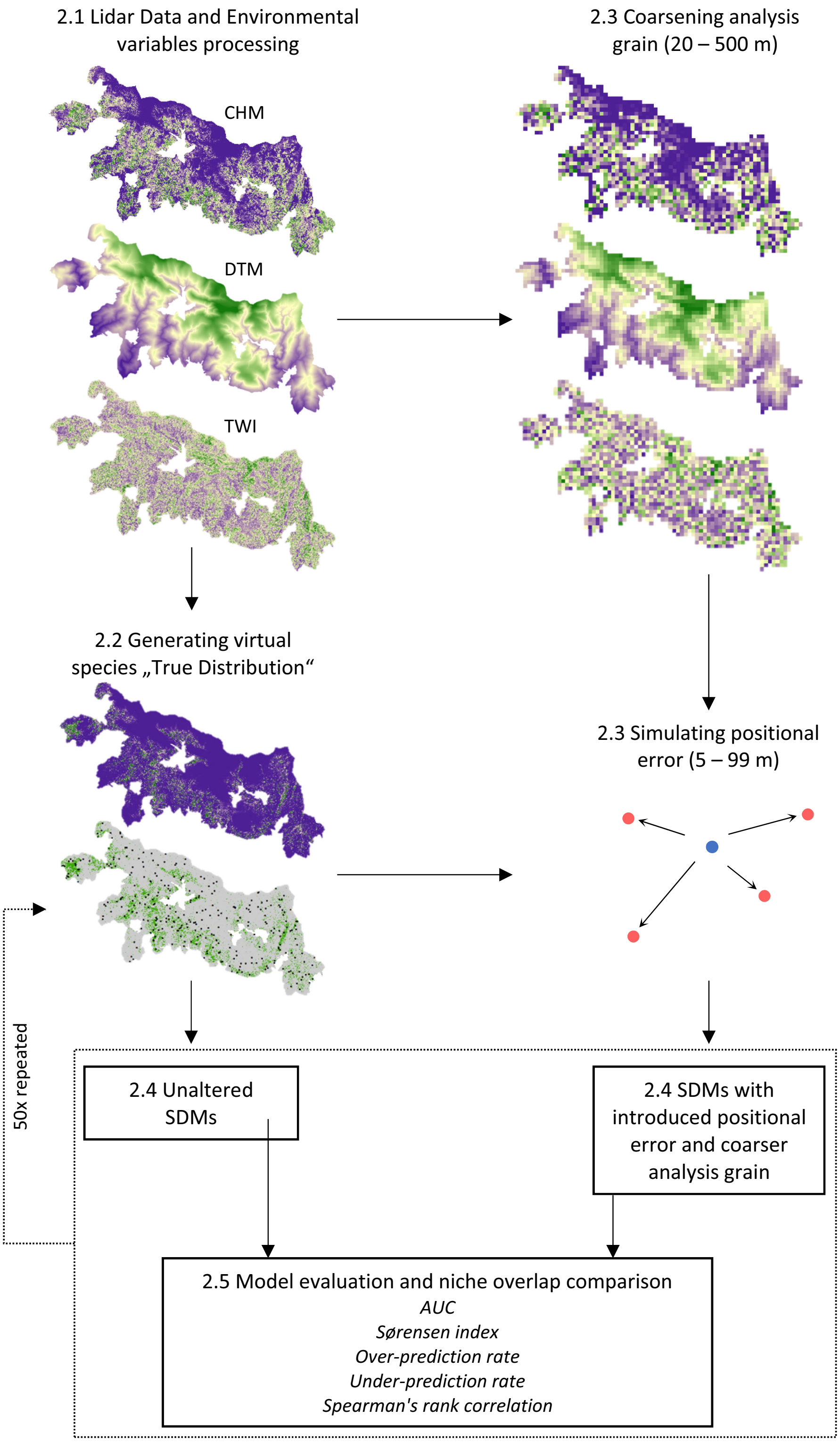

FIGURE 1 Overview of the modelling process. We first acquired and processed LiDAR data and selected three fine-scale environmental predictors (canopy height model, topographic wetness index, digital terrain model; Section 2.1). Furthermore, we generated virtual species (2.2), simulated positional error in species occurrences, and coarsened analysis grain (2.3). We modelled species distribution with unaltered data as well as with shifted occurrences at various analysis grain sizes (2.4). In the last step, we evaluated models and compared their performance (2.5).

rank correlation (Figure 3). Note that the observed trends were independent of the validation metric.

3.4 | Trade-off between positional error and analysis grain

Finally, and most importantly, our results clearly showed that coarsening the analysis grain cannot compensate for the effect of positional error (Figure 4). For each scenario positional error (S1–S6), we can observe that models with an analysis grain coarser than the initial grain (5 m) performed, at best, equally well, but never better than those with initial grain (i.e. response grain). In addition, models with a positional error of 20–29 m (S3) and higher perform almost equally well regardless of the analysis grain. This applies to all used performance metrics and Spearman's rank correlation used to assess the species niche overlap (Figure 4, Figure A4).

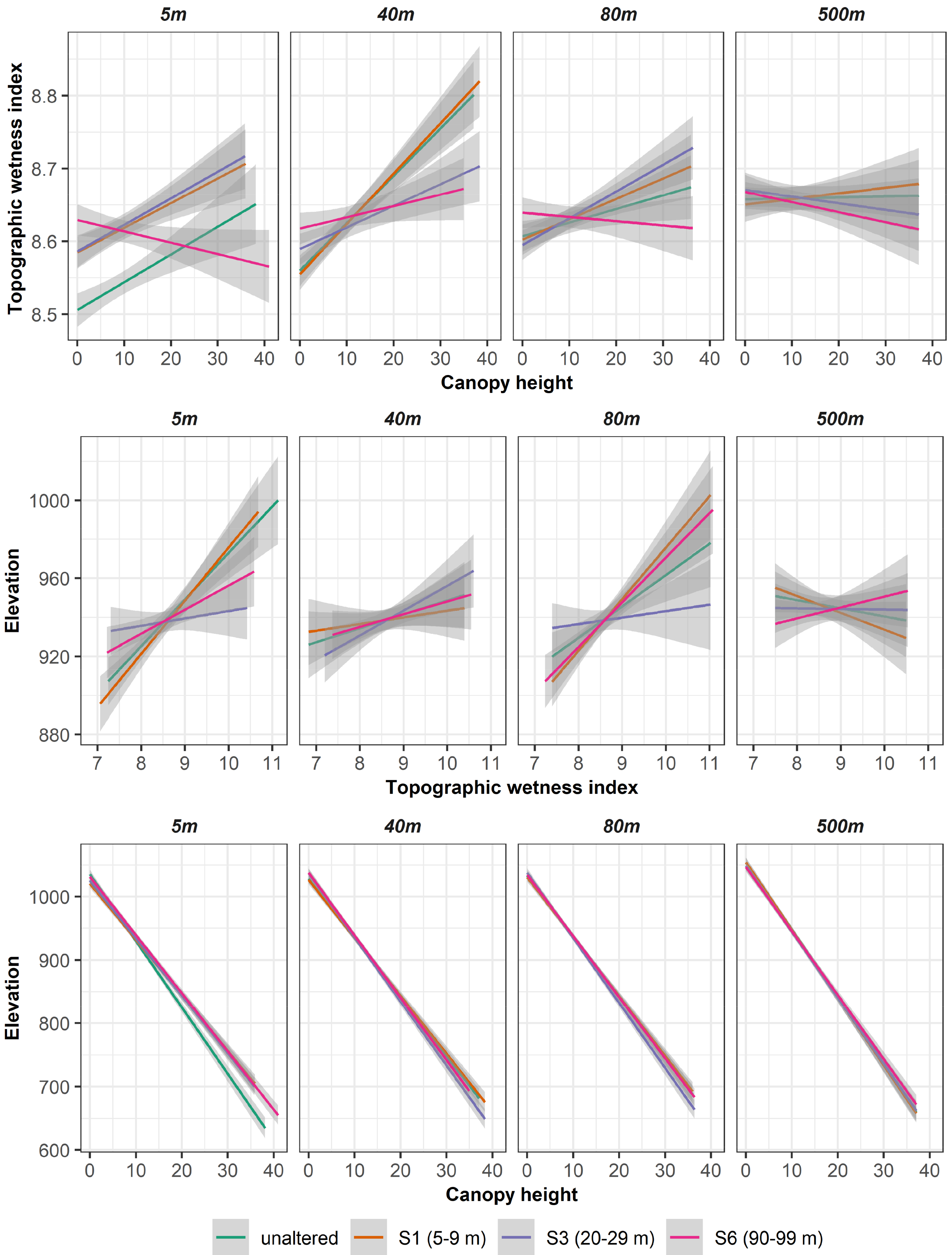

FIGURE 2 Comparison of changes in realized niche as a result of positional error in species occurrences and coarsening the analysis grain. Different colours show various levels of positional uncertainty while columns show different analysis grain. The line is obtained by linear regression and grey colour shows 95% confidence interval.

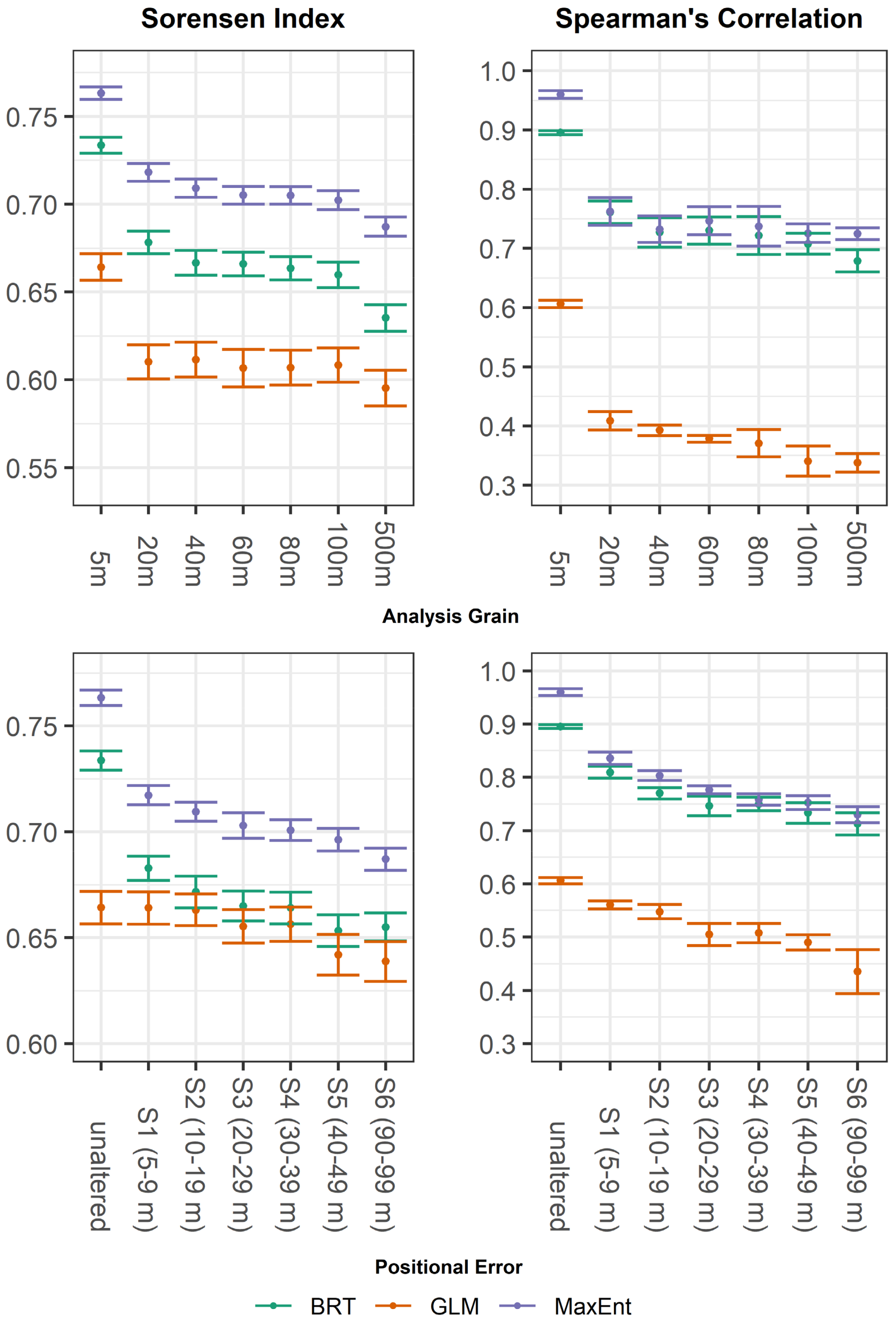

FIGURE 3 Sørensen index and Spearman's rank correlation scores of the different models. The first row shows results for models fitted with different analysis grains. The second row shows results for models fitted with an analysis grain of 5 m, but with positionally shifted species occurrences.

4 | DISCUSSION

In this study, we focused on the trade-off between the analysis grain and positional error in fine-scale SDMs. We simulated virtual species at 5 m resolution, coarsened the analysis grain (5–500 m) and introduced positional error (5–99 m) to evaluate their individual effects and potential trade-offs between them. Our results showed a negative effect of coarsening the analysis grain on SDMs performance. All modelling techniques were sensitive to the change in analysis grain (see also Guisan et al., 2007 for an analysis of the sensitivity of 10 modelling techniques to the change in grain size). Although this could be perceived as a negative, we believe that this is actually a positive characteristic, as it means that these models are sensitive to the use of an (in)appropriate resolution of the analysis grain. Similarly, introducing positional error led to a decrease in the discriminative ability of all modelling methods; yet, and importantly, coarsening the analysis grain did not offset for the effects of positional error.

The correct choice of the analysis grain is an important part of the overall modelling process and is affected by several other modelling choices. Ideally, the analysis grain is dictated by the species ecology and the objectives of the study, that is, it must match the response grain (Mertes & Jetz, 2018) but it could be also affected by sampling processes of species occurrences (Chase & Knight, 2013; Hurlbert & Jetz, 2007; Rahbek, 2005) and by the spatial extent of the study area. The spatial extent and resolution of the response variable govern what explanatory variables can be expected to act in determining species distribution (Pearson & Dawson, 2003).

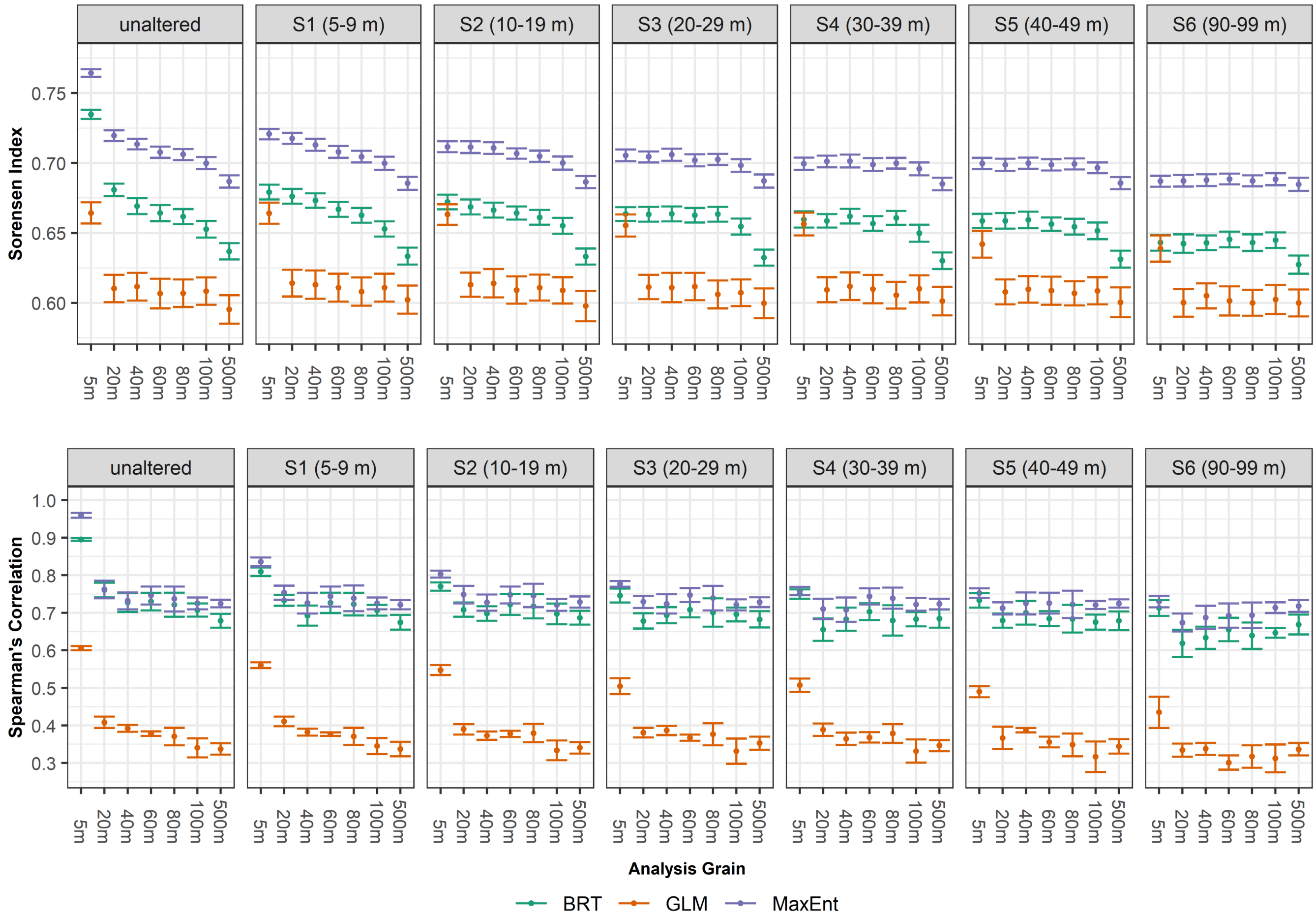

FIGURE 4 Sørensen index and Spearman's rank correlation scores according to different analysis grains and positional error scenarios (unaltered and S1–S6).

Typically, it is assumed that climate defines the distribution of species at very broad spatial scales (e.g. an extent of a whole continent and resolution of 100 km^2). At successively finer resolutions and at regional extents, topography or biotic interactions may become the most important factors controlling species occurrence, whereas at even finer resolutions, vegetation structure or presence of individual land cover categories (e.g. water bodies) can play a role (Gábor, Šímová, et al., 2022). However, some studies suggest that biotic interactions may shape species distribution across all spatial extents (Alexander et al., 2015; Wisz et al., 2013). Generally speaking, the importance of environmental factors varies with the adopted resolution and extent of the study and factors that are important at one resolution and extent can lose their importance at others (Corsi et al., 2000).

There are two typical situations regarding the choice of the analysis grain in species distribution modelling: (a) we know the response grain and have fine-scale data available or (b) we do not know the response grain and/or the analysis grain is chosen based on data availability rather than species ecology (Graf et al., 2005; Holland et al., 2004; Lechner et al., 2012; Martin & Fahrig, 2012; Mertes et al., 2020; Stuber & Fontaine, 2019). The first situation is represented in this study by the range of analysis grains from 5 m to 100 m, and the second by the 500 m grain. It should be noted that models are regularly built using an even coarser analysis grain than those tested in this study (e.g. 5 km or 10 km when using atlas data; Jetz et al., 2012). However, several studies have already tested the general effect of changing the grain of the response variable on modelling the species distribution in situations where the spatial resolution of the response variable was considerably coarser than the assumed response grain. For example, Seo et al. (2009) examined SDMs dynamics across a 64-fold (1 km to 64 km) change in the grain of the response variable and found that model performance decreased with increasing resolution. Similarly, Kaliontzopoulou et al. (2008) reported decreasing model performance at the 10 km response variable resolution compared with 1 km resolution.

Our results show that compensating position errors by coarsening the analysis grain does not lead to an improvement of the model performance in any of the scenarios investigated (Figure 4, A4). This is true even for very coarse analysis, that is, an analysis grain several orders of magnitude larger than the expected response grain. Therefore, based on our results and the results of the above-mentioned studies, we recommend using an analysis grain as fine as possible (or, in other words, as close to the response grain as possible), even if the available species occurrences suffer from positional error. This is consistent with recent findings by Mertes and Jetz (2018), who showed that coarsening the analysis grain can negatively affect intrinsic fine-scale heterogeneity in environmental variables (i.e. the pattern of spatial autocorrelation inherent in an environmental variable) and lead to variables that strongly influence distribution patterns being discarded simply because of their low explanatory power at such coarsened resolution. On the other hand, this contradicts the widely held assumption that coarsening the analysis grain can compensate for the negative effect of positional errors on model performance (Engler et al., 2004; Keil et al., 2014; Moudrý & Šímová, 2012; Sillero & Barbosa, 2021; Vollering et al., 2016), but this has never been thoroughly tested. Our results show that above a certain level of positional error (approximately five times higher than the response grain), models perform almost the same regardless of the analysis grain. Therefore, if there is considerable positional error in species occurrence data, users are unlikely to gain anything from making additional efforts to obtain higher-resolution data (but see Šímová et al., 2019) unless they also minimize the positional error.

Our findings and recommendations, however, do not mean that negative effects of the positional error can be ignored. On the contrary, the inability to compensate for the positional error by coarsening the analysis grain underscores the importance of careful georeferencing of species occurrence data. Our results show that the largest decrease in model performance occurs in the smallest simulated positional error (i.e. as soon as an error is introduced). This is consistent with previous studies and their conclusions that more accurate georeferencing approaches generally produce better performing SDMs (Gábor, Moudrý, Lecours, et al., 2020; Lash et al., 2012; Tulowiecki et al., 2015; Zhang et al., 2018). For example, Lash et al. (2012) have shown that using less accurate automated georeferencing methods is problematic in mapping the occurrence of monkeypox and modelling its transmission risk in Africa. The same limitations have been reported by Tulowiecki et al. (2015) for presettlement land survey records in North America that are useful for modelling the past distribution of tree species (e.g. Tulowiecki, 2020). On the other hand, it is fair to point out that Graham et al. (2008) concluded that SDMs are generally robust to positional errors. Similarly, Fernandez et al. (2009) concluded that while the models are somewhat sensitive to positional error, this sensitivity is considerably less than the sensitivity to the modelling method.

However, accurate georeferencing is an extremely time-consuming and labour-intensive process. In particular, georeferencing historical records can be challenging because in some parts of the world it is difficult to find suitable reference data with which to match place names. Guidelines for georeferencing exists (Wieczorek et al., 2004), and some heuristic approaches have been proposed to improve models created with poorly georeferenced data. These methods are applicable depending on the source of positional error and the available auxiliary data. For example, Hefley et al. (2014) used regression calibration to reduce the bias in coefficient estimates caused by the positional error. However, this approach requires that at least part of the data has locations recorded without error. Recently, Zhang et al. (2018) proposed a different approach to mitigate positional error in fine analysis grains (e.g. errors of tens of meters caused, for example, by the difference in position of the species and the observer). They narrowed down possible locations of species occurrences using auxiliary data such as the presence of habitat preferred by the species (e.g. forest), the assumed minimum and maximum distance (i.e. minimum distance the species keeps from the observer and the maximum distance at which the observer can see the species), and the observer's field of view (i.e. visibility analysis using a DTM; Lagner et al., 2018).

We intentionally developed our models with fine-scale environmental data that are increasingly adopted for SDMs (e.g. de Vries et al., 2021; Guillaume et al., 2021; Mitchell et al., 2017). Although so far, such data are typically used in models developed to assess species–environment relationships at a landscape scale, it has been highlighted that they can be crucial for understanding species distributions at global scales (Lembrechts, Lenoir, et al., 2019; Lembrechts, Nijs, & Lenoir, 2019; Stark & Fridley, 2022; Zellweger et al., 2019). Moreover, such fine-scale environmental data tend to be more heterogeneous, and hence species occurrences might easier end up in unsuitable environment, which can negatively affect SDMs (see Naimi et al., 2011, 2014). Therefore, understanding the interaction of analysis grain and positional error at fine-grain is crucial for future development of SDMs for conservation and climate change studies.

It is important to note that the effect of analysis grain and positional error is dependent on the magnitude of the potential change of the analysis grain (not the grain itself) and similarly the effect of positional error depends on the ratio between the magnitude of the positional error and the analysis grain. In addition, the magnitude of the effect will be affected by other characteristics. For example, it has been shown that the magnitude of the negative effect of positional error is related to species characteristics, such as niche (Gábor, Moudrý, Lecours, et al., 2020; Tulowiecki et al., 2015; Visscher, 2006) and heterogeneity in environmental variables (i.e. spatial autocorrelation; Naimi et al., 2011, 2014). For instance, models for species with relatively wide niche breadth and a region dominated by highly autocorrelated environmental variables or a single habitat will be relatively unaffected by positional error. On the contrary, the models for a region with abrupt changes (e.g. fragmented habitats) and for species with narrow niche breadth will be negatively affected with positional error in species data (see Naimi et al., 2011, 2014; Visscher, 2006). Therefore, our conclusions are also applicable into analysis using relatively coarse analysis grain, especially for SDMs developed for a region with abrupt changes in environment (e.g. fragmented habitats) and for species with narrow niche breadth (see Gábor, Moudrý, Barták, et al., 2020, Gábor, Moudrý, Lecours, et al. 2020; Naimi et al., 2011, 2014).

In this study, we examined how, in a species distribution modelling context, analysis grain and positional error in species occurrences interact. Our particular objective was to answer the question of whether the analysis grain is best kept close to the response grain or whether it should instead be coarsened to minimize the negative effects of positional errors in species occurrences on model performance, as suggested by several authors. We showed that a coarsened analysis grain is not able to compensate for the effects of positional errors. Thus, for data with unknown positional accuracy, we recommend keeping the analysis grain as close as possible to the response grain rather than coarsening it. We highlight that positional error in species occurrence cannot be overlooked and that great attention needs to be paid to the measurement and georeferencing techniques used to minimize positional error.

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: study conception and design: Lukáš Gábor, Vítězslav Moudrý, Walter Jetz; analysis and interpretation of results: Lukáš Gábor, Vítězslav Moudrý, Vojtěch Barták; draft manuscript preparation: Alejandra Zarzo-Arias, Anna Cord, Duccio Rocchini, Lukáš Gábor, Muyang Lu, Marco Malavasi, Vojtěch Barták, Vítězslav Moudrý, Walter Jetz; All authors reviewed the results and approved the final version of the manuscript.

ACKNOWLEDGEMENTS

The authors would like to thank the Krkonose Mountains National Park for providing LiDAR data. In addition, we would like to thank subject editor, associate editor and two anonymous reviewers for helpful feedback that allowed us to improve this article. As always, any remaining errors are our own. The first author would also like to recognize the neverending support from the Oden house family, which includes Fatima Leite, Andy Knight, Michael Glerup and Kate Glerup. Even though we were brought together by sheer coincidence, it is incredible to see how close we became in such a short period of time. I do not worry about when we will inevitably part ways physically, as I know you will continually remain influential in my life.

FUNDING INFORMATION

This research was funded by the Internal Grant Agency of Faculty of Environmental Sciences, Czech Univ. of Life Sciences Prague, grant no. 2020B0009 and by the Technological grant agency of the Czech Republic (grant no. SS02030018 DivLand). In addition, this paper was made possible by generous support from Fulbright-Masaryk program sponsored by U.S. and Czech governments, which provided Lukáš Gábor with the opportunity to conduct research at Yale University.

CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/2041-210X.13956.

DATA AVAILABILITY STATEMENT

The LiDAR data are owned by Krkonose Mountains National Park and were provided to the authors from the Czech University of Life Sciences Prague for scientific purposes. However, the authors are unable to share the data as they are not the owners. The LiDAR data are available upon request to the Department of Informatics and GIS for research purposes (https://www.krnap.cz/adresar/?firstname=&surname=&department=Pracovi%C5%A1t%C4%9B+informatiky+a+GIS&filter=1). The 'vs' object and R script with sampling and shifting occurrence data are available via the Dryad repository (Gábor, Jetz, et al., 2022, https://doi.org/10.5061/dryad.79cnp5hx3).

ORCID

Lukáš Gábor https://orcid.org/0000-0001-6137-0994
Walter Jetz https://orcid.org/0000-0002-1971-7277
Muyang Lu https://orcid.org/0000-0002-4949-8837
Duccio Rocchini https://orcid.org/0000-0003-0087-0594
Anna Cord https://orcid.org/0000-0003-3183-8482
Marco Malavasi https://orcid.org/0000-0002-9639-1784
Alejandra Zarzo-Arias https://orcid.org/0000-0001-5496-0144
Vojtěch Barták https://orcid.org/0000-0001-9887-1290
Vítězslav Moudrý https://orcid.org/0000-0002-3194-451X

REFERENCES

Alexander, J. M., Diez, J. M., & Levine, J. M. (2015). Novel competitors shape species' responses to climate change. *Nature*, *525*, 515–518.

Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, *43*(6), 1223–1232.

Anderson, R. P., Araújo, M., Guisan, A., Lobo, J. M., Martínez-Meyer, E., Peterson, A. T., & Soberón, J. (2016). *Final report of the task group on GBIF data fitness for use in distribution modelling*. Global Biodiversity Information Facility.

Araújo, M. B., Anderson, R. P., Barbosa, A. M., 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.

Bakx, T. R., Koma, Z., Seijmonsbergen, A. C., & Kissling, W. D. (2019). Use and categorization of light detection and ranging vegetation metrics in avian diversity and species distribution research. *Diversity and Distributions*, *25*(7), 1045–1059.

Ballesteros-Mejia, L., Kitching, I. J., Jetz, W., & Beck, J. (2017). Putting insects on the map: Near-global variation in sphingid moth richness along spatial and environmental gradients. *Ecography*, *40*(6), 698–708.

Barbosa, M. A., Real, R., Muñoz, A. R., & Brown, J. A. (2013). New measures for assessing model equilibrium and prediction mismatch in species distribution models. *Diversity and Distributions*, *19*(10), 1333–1338.

Chase, J. M., & Knight, T. M. (2013). Scale-dependent effect sizes of ecological drivers on biodiversity: Why standardised sampling is not enough. *Ecology Letters*, *16*, 17–26.

Cheng, Y., Tjaden, N. B., Jaeschke, A., Thomas, S. M., & Beierkuhnlein, C. (2021). Using centroids of spatial units in ecological niche modelling: Effects on model performance in the context of environmental data grain size. *Global Ecology and Biogeography*, *30*(3), 611–621.

Corsi, F., De Leeuw, J., & Skidmore, A. (2000). Modeling species distribution with GIS. In L. Boitani & T. K. Fuller (Eds.), *Research techniques in animal ecology: Controversies and consequences* (pp. 389–434). Columbia University Press.

de Vries, J. P. R., Koma, Z., WallisDeVries, M. F., & Kissling, W. D. (2021). Identifying fine-scale habitat preferences of threatened butterflies using airborne laser scanning. *Diversity and Distributions*, *27*, 1251–1264.

Dungan, J. L., Perry, J. N., Dale, M. R. T., Legendre, P., Citron-Pousty, S., Fortin, M. J., Jakomulska, A., Miriti, M., & Rosenberg, M. (2002). A balanced view of scale in spatial statistical analysis. *Ecography*, *25*(5), 626–640.

Ellis-Soto, D., Merow, C., Amatulli, G., Parra, J. L., & Jetz, W. (2021). Continental-scale 1 km hummingbird diversity derived from fusing point records with lateral and elevational expert information. *Ecography*, *44*(4), 640–652.

Engler, R., Guisan, A., & Rechsteiner, L. (2004). An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Journal of Applied Ecology*, *41*(2), 263–274.

Fernandez, M., Blum, S., Reichle, S., Guo, Q., Holzman, B., & Hamilton, H. (2009). Locality uncertainty and the differential performance of four common niche-based modeling techniques. *Biodiversity Informatics*, *6*, 36–52.

Ferrier, S., Jetz, W., & Scharlemann, J. (2017). Biodiversity modelling as part of an observation system. In *The GEO handbook on biodiversity observation networks* (pp. 239–257). Springer.

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*, 38–49.

Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: A statistical view of boosting (with discussion and a rejoinder by the authors). *Annals of Statistics*, *28*(2), 337–407.

Gábor, L., Jetz, W., Lu, M., Rocchini, D., Cord, A., Malavasi, M., Zarzo-Arias, A., Barták, V., & Moudrý, V. (2022). Data from: Positional errors in species distribution modelling are not overcome by the coarser grains of analysis, *Dryad Digital Repository*, https://doi.org/10.5061/dryad.79cnp5hx3

Gábor, L., Moudrý, V., Barták, V., & Lecours, V. (2020). How do species and data characteristics affect species distribution models and when to use environmental filtering? *International Journal of Geographical Information Science*, *34*(8), 1567–1584.

Gábor, L., Moudrý, V., Lecours, V., Malavasi, M., Barták, V., Fogl, M., Šímová, P., Rocchini, D., & Václavík, T. (2020). The effect of positional error on fine scale species distribution models increases for specialist species. *Ecography*, *43*(2), 256–269.

Gábor, L., Šímová, P., Keil, P., Zarzo-Arias, A., Marsh, C. J., Rocchini, D., Malavasi, M., Bartík, V., & Moudrý, V. (2022). Habitats as predictors in species distribution models: Shall we use continuous or binary data? *Ecography*, *e06022*.

Graf, R. F., Bollmann, K., Suter, W., & Bugmann, H. (2005). The importance of spatial scale in habitat models: Capercaillie in the Swiss Alps. *Landscape Ecology*, *20*(6), 703–717.

Graham, C. H., Elith, J., Hijmans, R. J., Guisan, A., Townsend Peterson, A., Loiselle, B. A., & NCEAS Predicting Species Distributions Working Group. (2008). The influence of spatial errors in species occurrence data used in distribution models. *Journal of Applied Ecology*, *45*(1), 239–247.

Grimmett, L., Whitsed, R., & Horta, A. (2021). Creating virtual species to test species distribution models: The importance of landscape structure, dispersal and population processes. *Ecography*, *44*(5), 753–765.

Gueta, T., & Carmel, Y. (2016). Quantifying the value of user-level data cleaning for big data: A case study using mammal distribution models. *Ecological Informatics*, *34*, 139–145.

Guillaume, A. S., Leempoel, K., Rochat, E., Rogivue, A., Kasser, M., Gugerli, F., Parisod, C. H., & Joost, S. (2021). Multiscale very high resolution topographic models in alpine ecology: Pros and cons of airborne LiDAR and drone-based stereo-photogrammetry technologies. *Remote Sensing*, *13*(8), 1588.

Guisan, A., Graham, C. H., Elith, J., Huettmann, F., & NCEAS Species Distribution Modelling Group. (2007). Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions*, *13*(3), 332–340.

Haesen, S., Lembrechts, J. J., De Frenne, P., Lenoir, J., Aalto, J., Ashcroft, M. B., Kopecký, M., Luoto, M., Maclean, I., Nijs, I., Niittynen, P., van den Hoogen, J., Arriga, N., Brůna, J., Buchmann, N., Čiliak, M., Collalti, A., de Lombaerde, E., Descombes, P., … Van Meerbeek, K. (2021). ForestTemp–Sub-canopy microclimate temperatures of European forests. *Global Change Biology*, *27*(23), 6307–6319.

Hallman, T. A., & Robinson, W. D. (2020). Deciphering ecology from statistical artefacts: Competing influence of sample size, prevalence and habitat specialization on species distribution models and how small evaluation datasets can inflate metrics of performance. *Diversity and Distributions*, *26*(3), 315–328.

Hefley, T. J., Baasch, D. M., Tyre, A. J., & Blankenship, E. E. (2014). Correction of location errors for presence-only species distribution models. *Methods in Ecology and Evolution*, *5*(3), 207–214.

Holland, J. D., Bert, D. G., & Fahrig, L. (2004). Determining the spatial scale of species' response to habitat. *Bioscience*, *54*(3), 227–233.

Hurlbert, A. H., & Jetz, W. (2007). Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(33), 13384–13389.

Inman, R., Franklin, J., Esque, T., & Nussear, K. (2021). Comparing sample bias correction methods for species distribution modeling using virtual species. *Ecosphere*, *12*(3), e03422.

Jetz, W., McPherson, J. M., & Guralnick, R. P. (2012). Integrating biodiversity distribution knowledge: Toward a global map of life. *Trends in Ecology & Evolution*, *27*(3), 151–159.

Jiménez-Valverde, A. (2012). Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Global Ecology and Biogeography*, *21*(4), 498–507.

Jiménez-Valverde, A. (2020). Sample size for the evaluation of presence-absence models. *Ecological Indicators*, *114*, 106289.

Jiménez-Valverde, A., Lobo, J. M., & Hortal, J. (2009). The effect of prevalence and its interaction with sample size on the reliability of species distribution models. *Community Ecology*, *10*(2), 196–205.

Johnson, C. J., & Gillingham, M. P. (2008). Sensitivity of species-distribution models to error, bias, and model design: An application to resource selection functions for woodland caribou. *Ecological Modelling*, *213*(2), 143–155.

Kaliontzopoulou, A., Brito, J. C., Carretero, M. A., Larbes, S., & Harris, D. J. (2008). Modelling the partially unknown distribution of wall lizards (Podarcis) in North Africa: Ecological affinities, potential areas of occurrence, and methodological constraints. *Canadian Journal of Zoology*, *86*(9), 992–1001.

Keil, P., Wilson, A. M., & Jetz, W. (2014). Uncertainty, priors, autocorrelation and disparate data in downscaling of species distributions. *Diversity and Distributions*, *20*(7), 797–812.

Khosravipour, A., Skidmore, A. K., & Isenburg, M. (2016). Generating spike-free digital surface models using LiDAR raw point clouds: A new approach for forestry applications. *International Journal of Applied Earth Observation and Geoinformation*, *52*, 104–114.

Klápště, P., Fogl, M., Barták, V., Gdulová, K., Urban, R., & Moudrý, V. (2020). Sensitivity analysis of parameters and contrasting performance of ground filtering algorithms with UAV photogrammetry-based and LiDAR point clouds. *International Journal of Digital Earth*, *13*(12), 1672–1694.

Lagner, O., Klouček, T., & Šímová, P. (2018). Impact of input data (in) accuracy on overestimation of visible area in digital viewshed models. *PeerJ*, *6*, e4835.

Lash, R. R., Carroll, D. S., Hughes, C. M., Nakazawa, Y., Karem, K., Damon, I. K., & Peterson, A. T. (2012). Effects of georeferencing effort on mapping monkeypox case distributions and transmission risk. *International Journal of Health Geographics*, *11*(1), 1–12.

Lechner, A. M., Langford, W. T., Jones, S. D., Bekessy, S. A., & Gordon, A. (2012). Investigating species–environment relationships at multiple scales: Differentiating between intrinsic scale and the modifiable areal unit problem. *Ecological Complexity*, *11*, 91–102.

Lecours, V., Devillers, R., Schneider, D. C., Lucieer, V. L., Brown, C. J., & Edinger, E. N. (2015). Spatial scale and geographic context in benthic habitat mapping: Review and future directions. *Marine Ecology Progress Series*, *535*, 259–284.

Lecours, V., Gábor, L., Edinger, E., & Devillers, R. (2020). Fine-scale habitat characterization of The Gully, the Flemish Cap, and the Orphan Knoll, Northwest Atlantic, with a focus on cold-water corals. In *Seafloor geomorphology as benthic habitat* (pp. 735–751). Elsevier.

Lembrechts, J. J., Lenoir, J., Roth, N., Hattab, T., Milbau, A., Haider, S., Pellissier, L., Pauchard, A., Backes, A. R., Dimarco, R. D., Nunez, M. A., Aalto, J., & Nijs, I. (2019). Comparing temperature data sources for use in species distribution models: From in-situ logging to remote sensing. *Global Ecology and Biogeography*, *28*(11), 1578–1596.

Lembrechts, J. J., Nijs, I., & Lenoir, J. (2019). Incorporating microclimate into species distribution models. *Ecography*, *42*(7), 1267–1279.

Leroy, B., Delsol, R., Hugueny, 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.

Leroy, B., Meynard, C. N., Bellard, C., & Courchamp, F. (2016). Virtualspecies, an R package to generate virtual species distributions. *Ecography*, *39*(6), 599–607.

Li, R., Ranipeta, A., Wilshire, J., Malczyk, J., Duong, M., Guralnick, R., Wilson, A., & Jetz, W. (2021). A cloud-based toolbox for the versatile environmental annotation of biodiversity data. *PLoS Biology*, *19*(11), e3001460.

Li, W., & Guo, Q. (2013). How to assess the prediction accuracy of species presence–absence models without absence data? *Ecography*, *36*(7), 788–799.

Liu, C., Newell, G., & White, M. (2019). The effect of sample size on the accuracy of species distribution models: Considering both presences and pseudo-absences or background sites. *Ecography*, *42*, 535–548.

Lobo, J. M., Jiménez-Valverde, A., & Real, R. (2008). AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, *17*(2), 145–151.

MacKinnon, J. G., & White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, *29*(3), 305–325.

Martin, A. E., & Fahrig, L. (2012). Measuring and selecting scales of effect for landscape predictors in species–habitat models. *Ecological Applications*, *22*(8), 2277–2292.

Mayor, S. J., Schneider, D. C., Schaefer, J. A., & Mahoney, S. P. (2009). Habitat selection at multiple scales. *Ecoscience*, *16*(2), 238–247.

McGarigal, K., Wan, H. Y., Zeller, K. A., Timm, B. C., & Cushman, S. A. (2016). Multi-scale habitat selection modeling: A review and outlook. *Landscape Ecology*, *31*(6), 1161–1175.

McPherson, J. M., Jetz, W., & Rogers, D. J. (2004). The effects of species' range sizes on the accuracy of distribution models: Ecological phenomenon or statistical artefact? *Journal of Applied Ecology*, *41*(5), 811–823.

McPherson, M. J., & Jetz, W. (2007). Effects of species' ecology on the accuracy of distribution models. *Ecography*, *30*(1), 135–151.

Merow, C., Maitner, B. S., Owens, H. L., Kass, J. M., Enquist, B. J., Jetz, W., & Guralnick, R. (2019). Species' range model metadata standards: RMMS. *Global Ecology and Biogeography*, *28*(12), 1912–1924.

Mertes, K., Jarzyna, M. A., & Jetz, W. (2020). Hierarchical multi-grain models improve descriptions of species' environmental associations, distribution, and abundance. *Ecological Applications*, *30*(6), e02117.

Mertes, K., & Jetz, W. (2018). Disentangling scale dependencies in species environmental niches and distributions. *Ecography*, *41*(10), 1604–1615.

Meynard, C. N., Leroy, B., & Kaplan, D. M. (2019). Testing methods in species distribution modelling using virtual species: What have we learnt and what are we missing? *Ecography*, *42*(12), 2021–2036.

Miguet, P., Jackson, H. B., Jackson, N. D., Martin, A. E., & Fahrig, L. (2016). What determines the spatial extent of landscape effects on species? *Landscape Ecology*, *31*(6), 1177–1194.

Mitchell, P. J., Monk, J., & Laurenson, L. (2017). Sensitivity of fine-scale species distribution models to locational uncertainty in occurrence data across multiple sample sizes. *Methods in Ecology and Evolution*, *8*(1), 12–21.

Montgomery, R. A., Roloff, G. J., & Hoef, J. M. V. (2011). Implications of ignoring telemetry error on inference in wildlife resource use models. *The Journal of Wildlife Management*, *75*(3), 702–708.

Moudrý, V. (2015). Modelling species distributions with simulated virtual species. *Journal of Biogeography*, *42*(8), 1365–1366.

Moudrý, V., & Devillers, R. (2020). Quality and usability challenges of global marine biodiversity databases: An example for marine mammal data. *Ecological Informatics*, *56*, 101051.

Moudrý, V., Lecours, V., Malavasi, M., Misiuk, B., Gábor, L., Gdulová, K., Šímová, P., & Wild, J. (2019). Potential pitfalls in rescaling digital terrain model-derived attributes for ecological studies. *Ecological Informatics*, *54*, 100987.

Moudrý, V., Moudrá, L., Barták, V., Bejček, V., Gdulová, K., Hendrychová, M., Moravec, D., Musil, P., Rocchini, D., Šťastný, K., Volf, O., & Šálek, M. (2021). The role of the vegetation structure, primary productivity and senescence derived from airborne LiDAR and hyperspectral data for birds diversity and rarity on a restored site. *Landscape and Urban Planning*, *210*, 104064.

Moudrý, V., & Šímová, P. (2012). Influence of positional accuracy, sample size and scale on modelling species distributions: A review. *International Journal of Geographical Information Science*, *26*(11), 2083–2095.

Naimi, B., Hamm, N. A., Groen, T. A., Skidmore, A. K., & Toxopeus, A. G. (2014). Where is positional uncertainty a problem for species distribution modelling? *Ecography*, *37*(2), 191–203.

Naimi, B., Skidmore, A. K., Groen, T. A., & Hamm, N. A. (2011). Spatial autocorrelation in predictors reduces the impact of positional uncertainty in occurrence data on species distribution modelling. *Journal of Biogeography*, *38*(8), 1497–1509.

Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, *135*(3), 370–384.

Oksanen, J., & Minchin, P. R. (2002). Continuum theory revisited: What shape are species responses along ecological gradients? *Ecological Modelling*, *157*(2–3), 119–129.

Osborne, P. E., & Leitão, P. J. (2009). Effects of species and habitat positional errors on the performance and interpretation of species distribution models. *Diversity and Distributions*, *15*(4), 671–681.

Pearson, R. G., & Dawson, T. P. (2003). Predicting the impacts of climate change on the distribution of species: Are bioclimate envelope models useful? *Global Ecology and Biogeography*, *12*(5), 361–371.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*(3–4), 231–259.

Phillips S. J., Dudík, M., & Schapire, R. E. (2020). *[Internet] Maxent software for modeling species niches and distributions (version 3.4.3).* Retrieved from http://biodiversityinformatics.amnh.org/open_source/maxent/

Pradervand, J. N., Dubuis, A., Pellissier, L., Guisan, A., & Randin, C. (2014). Very high resolution environmental predictors in species distribution models: Moving beyond topography? *Progress in Physical Geography*, *38*(1), 79–96.

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rahbek, C. (2005). The role of spatial scale and the perception of large-scale species-richness patterns. *Ecology Letters*, *8*(2), 224–239.

Reif, J., Reifová, R., Skoracka, A., & Kuczyński, L. (2018). Competition-driven niche segregation on a landscape scale: Evidence for escaping from syntopy towards allotopy in two coexisting sibling passerine species. *Journal of Animal Ecology*, *87*(3), 774–789.

Robertson, M. P., Visser, V., & Hui, C. (2016). Biogeo: An R package for assessing and improving data quality of occurrence record datasets. *Ecography*, *39*(4), 394–401.

Rödder, D., & Engler, J. O. (2011). Quantitative metrics of overlaps in Grinnellian niches: Advances and possible drawbacks. *Global Ecology and Biogeography*, *20*(6), 915–927.

Schneider, D. C. (2001). The rise of the concept of scale in ecology: The concept of scale is evolving from verbal expression to quantitative expression. *Bioscience*, *51*(7), 545–553.

Seo, C., Thorne, J. H., Hannah, L., & Thuiller, W. (2009). Scale effects in species distribution models: Implications for conservation planning under climate change. *Biology Letters*, *5*(1), 39–43.

Sillero, N., & Barbosa, A. M. (2021). Common mistakes in ecological niche models. *International Journal of Geographical Information Science*, *35*(2), 213–226.

Sillero, N., & Goncalves-Seco, L. (2014). Spatial structure analysis of a reptile community with airborne LiDAR data. *International Journal of Geographical Information Science*, *28*(8), 1709–1722.

Simoes, M., Romero-Alvarez, D., Nuñez-Penichet, C., Jiménez, L., & Cobos, M. E. (2020). General theory and good practices in ecological niche modeling: A basic guide. *Biodiversity Informatics*, *15*(2), 67–68.

Simonson, W. D., Allen, H. D., & Coomes, D. A. (2014). Applications of airborne lidar for the assessment of animal species diversity. *Methods in Ecology and Evolution*, *5*(8), 719–729.

Šímová, P., Moudrý, V., Komárek, J., Hrach, K., & Fortin, M. J. (2019). Fine scale waterbody data improve prediction of waterbird occurrence despite coarse species data. *Ecography*, *42*(3), 511–520.

Stark, J. R., & Fridley, J. D. (2022). Microclimate-based species distribution models in complex forested terrain indicate widespread cryptic refugia under climate change. *Global Ecology and Biogeography*, *31*(3), 562–575.

Stuber, E. F., & Fontaine, J. J. (2019). How characteristic is the species characteristic selection scale? *Global Ecology and Biogeography*, *28*(12), 1839–1854.

Tulowiecki, S. J. (2020). Modeling the historical distribution of American chestnut (*Castanea dentata*) for potential restoration in western New York State, US. *Forest Ecology and Management*, *462*, 118003.

Tulowiecki, S. J., Larsen, C. P., & Wang, Y. C. (2015). Effects of positional error on modeling species distributions: A perspective using presettlement land survey records. *Plant Ecology*, *216*(1), 67–85.

van Proosdij, A. S., Sosef, M. S., Wieringa, J. J., & Raes, N. (2016). Minimum required number of specimen records to develop accurate species distribution models. *Ecography*, *39*, 542–552.

Visscher, R. D. (2006). GPS measurement error and resource selection functions in a fragmented landscape. *Ecography*, *29*(3), 458–464.

Vogeler, J. C., Hudak, A. T., Vierling, L. A., Evans, J., Green, P., & Vierling, K. T. (2014). Terrain and vegetation structural influences on local avian species richness in two mixed-conifer forests. *Remote Sensing of Environment*, *147*, 13–22.

Vollering, J., Schuiteman, A., de Vogel, E., van Vugt, R., & Raes, N. (2016). Phytogeography of New Guinean orchids: Patterns of species richness and turnover. *Journal of Biogeography*, *43*(1), 204–214.

Watcharamongkol, T., Christin, P. A., & Osborne, C. P. (2018). C4 photosynthesis evolved in warm climates but promoted migration to cooler ones. *Ecology Letters*, *21*(3), 376–383.

Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T., & Vieglais, D. (2012). Darwin Core: An evolving community-developed biodiversity data standard. *PLoS One*, *7*(1), e29715.

Wieczorek, J., Guo, Q., & Hijmans, R. (2004). The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *International Journal of Geographical Information Science*, *18*(8), 745–767.

Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H., Guisan, A., & NCEAS Predicting Species Distributions Working Group. (2008). Effects of sample size on the performance of species distribution models. *Diversity and distributions.*, *14*, 763–773.

Wisz, M. S., Pottier, J., Kissling, W. D., Pellissier, L., Lenoir, J., Damgaard, C. F., Dormann, C. F., Forchhammer, M. C., Grytnes, J. A., Guisan, A., Heikkinen, R. K., Hoye, T. T., Kuhn, I., Luoto, M., Maiorano, L., Nilsson, M. C. H., Normand, S., Ockinger, E., Schmidt, N. M., … Svenning, J. C. (2013). The role of biotic interactions in shaping distributions and realised assemblages of species: Implications for species distribution modelling. *Biological Reviews*, *88*(1), 15–30.

Wu, J., & Li, H. (2006). Concepts of scale and scaling. In *Scaling and uncertainty analysis in ecology* (pp. 3–15). Springer.

Wüest, R. O., Zimmermann, N. E., Zurell, D., Alexander, J. M., Fritz, S. A., Hof, C., Kreft, H., Normand, S., Cabral, J. S., Szekely, E., Thuiller, W., WIkelski, M., & Karger, D. N. (2020). Macroecology in the age of Big Data – Where to go from here? *Journal of Biogeography*, *47*(1), 1–12.

Zellweger, F., De Frenne, P., Lenoir, J., Rocchini, D., & Coomes, D. (2019). Advances in microclimate ecology arising from remote sensing. *Trends in Ecology & Evolution*, *34*(4), 327–341.

Zhang, G., Zhu, A. X., Huang, Z. P., & Xiao, W. (2018). A heuristic-based approach to mitigating positional errors in patrol data for species distribution modeling. *Transactions in GIS*, *22*(1), 202–216.

Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Duarte Ritter, C., Edler, D., Farooq, H., Herdean, A., Ariza, M., Scharn, R., Svantesson, S., Wengstrom, N., Zizka, V., & Antonelli, A. (2019). CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution*, *10*(5), 744–751.

Zurell, D., Berger, U., Cabral, J. S., Jeltsch, F., Meynard, C. N., Münkemüller, T., Nehrbass, N., Pagel, J., Reineking, B., Schroder, B., & Grimm, V. (2010). The virtual ecologist approach: Simulating data and observers. *Oikos*, *119*(4), 622–635.

SUPPORTING INFORMATION

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

How to cite this article: Gábor, L., Jetz, W., Lu, M., Rocchini, D., Cord, A., Malavasi, M., Zarzo-Arias, A., Barták, V., & Moudrý, V. (2022). Positional errors in species distribution modelling are not overcome by the coarser grains of analysis. *Methods in Ecology and Evolution*, *13*, 2289–2302. https://doi.org/10.1111/2041-210X.13956