Check for updates

Received: 28 July 2024 | Revised: 1 January 2025 | Accepted: 16 January 2025

DOI: 10.1111/cobi.70050

Conservation Biology

CONTRIBUTED PAPER

Impact of obscured data on species distribution models

Kyo Soung Koo[1,2] | **Ko-Huan Lee**[2,3] | **Dawon Lee**[4] | **Yikweon Jang**[5]

[1]Korean Environmental Geography Institute, Sejong, Republic of Korea

[2]Research Institute of EcoScience, Ewha Womans University, Seoul, Republic of Korea

[3]Commonwealth Scientific and Industrial Research Organisation (CSIRO), Health and Biosecurity, Canberra, Australian Capital Territory, Australia

[4]Department of Life Sciences, Ewha Womans University, Seoul, Republic of Korea

[5]Division of EcoScience, Ewha Womans University, Seoul, Republic of Korea

Correspondence
Kyo Soung Koo, Korean Environmental Geography Institute, Sejong 30141, Republic of Korea. Email: flqpfj85@gmail.com.

Yikweon Jang, Division of EcoScience, Ewha Womans University, Seoul 03760, Republic of Korea. Email: jangy@ewha.ac.kr

Article impact statement: Obscured data significantly distort the basic information and analysis results of species used for conservation and policy.

Funding information
Rural Development Administration, Grant/Award Number: RS-2024-00397542; Korea Environmental Industry & Technology Institute (KEITI), Grant/Award Number: 2021-002270789

Abstract

The lack of knowledge about geographic distribution and environmental preference can hinder conservation efforts for rare and threatened species. Open-source databases provide an opportunity to address these knowledge gaps through the geographic information they hold on species worldwide. However, to protect rare and endangered species, open-source databases often assign locations that do not match the original locations, which introduce inaccuracies in occurrence records (e.g., the "obscured" function in iNaturalist replaces the original location with a random location in a 0.2 × 0.2° cell). We tested the efficacy of the iNaturalist's obscured function in concealing geographic information and the function's impact on the species distribution modeling of 3 endangered species in South Korea: gold-spotted pond frogs (*Pelophylax chosenicus*), Reeves' turtles (*Mauremys reevesii*), and Mongolia racerunner (*Eremias argus*). We collected occurrence data (orginal data) for these 3 species and uploaded the data to iNaturalist. We then compared location, elevation, and habitat area in the original data set with these data in the obscured data set. To investigate the differences in species distribution, we ran species distribution models with both data sets. We also assessed the awareness of obscured function in peer-reviewed articles for which occurrence records from iNaturalist were used. The locations assigned by the obscured function significantly altered the geographic information of the species, including elevational range, habitat type, and environmental variables relevant to species distribution. Potential distributions estimated using locations assigned under the obscured function were different from those estimated using the original data. Only 4 out of 170 peer-reviewed articles acknowledged the presence of obscured data in iNaturalist, suggesting that most researchers are unaware of this issue. The locations assigned by the obscured function can cause serious problems in species distribution modeling and thus may negatively affect conservation of endangered species. We encourage researchers to thoroughly vet data obtained from open-source databases and urge database platforms to make it clear when data have been obscured.

KEYWORDS
biodiversity data, ecological model, GBIF, geoprivacy status, positional inaccuracy

INTRODUCTION

Open-source databases are popular platforms for storing biodiversity data and have accumulated billions of data points over the past decade (https://www.gbif.org/). This rapid development is fueled by advances in internet technology and, most importantly, the rise of citizen science worldwide (Bisby, 2000; Poisson et al., 2020). Notable open-source biodiversity databases include the Global Biodiversity Information Facility (GBIF), eBird, and iNaturalist (Lane & Edwards, 2007; Nugent, 2018; Sullivan et al., 2009). These databases play a major role in recording species distributions, identifying potential invasions, and facilitating policy decision-making processes (Ball-Damerow et al., 2019). Open-source data are particularly useful for conducting large-scale studies, such as predicting changes in species distribution caused by environmental and

Kyo Soung Koo and Ko-Huan Lee contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2025 The Author(s). *Conservation Biology* published by Wiley Periodicals LLC on behalf of Society for *Conservation Biology*.

climate changes and supporting conservation efforts for endangered species (Koldasbayeva et al., 2022; Pimm et al., 2015). The widespread use of these databases helps address information shortfalls in biodiversity, filling data gaps in geographical distribution (Wallacean shortfall) and environmental preferences (Hutchinsonian shortfall) (Hochkirch et al., 2021; Hortal et al., 2015).

Although data provided by citizen scientists can be prone to biased sampling or species misidentification (Fisher et al., 2022; Tang et al., 2021), most databases either employ experts or rely on experienced volunteers to review the data submitted by the public (Campbell et al., 2023; Rowley & Callaghan, 2020). This review process helps maintain the quality of the open-source data and reduces errors that may lead to inaccurate results in subsequent analyses (Costello et al., 2013, 2014). However, users' unfamiliarity with the database interfaces can also lead to errors in the exported data sets; these errors affect subsequent analyses. For example, misunderstandings regarding the max_target_seqs parameter in the NCBI BLAST database affected searches for best-fit sequences conducted for hundreds of scientific papers (Shah et al., 2019). Additionally, little is known about the extent to which the misuse of open-source databases alters the results of ecological modeling or how widespread this problem is in literature.

Open-source databases often include options to conceal the geographic information of rare or endangered species to prevent poaching, habitat destruction by nature enthusiasts, and conflicts between researchers and landowners (Lindenmayer & Scheele, 2017). iNaturalist, with over 220 million observations, uses functions called "private" and "obscured" to restrict public access to the geographic information of observations (iNaturalist, 2023). The private function completely hides the geographic information (i.e., coordinates and locality notes) from the public, and the obscured function replaces the original coordinates with a random location within a 0.2 × 0.2° cell (~500 km^2 at the equator) (Figure 1). The location of threatened or endangered species (based on IUCN red list) is concealed by default in iNaturalist's system. However, data uploaders can apply the private or obscured functions to nonthreatened or nonendangered species if they choose.

Users can obtain the coordinates of a randomly generated location (hereafter called the obscured location) by clicking the details tab in the iNaturalist interface (Figure 1). If multiple observations are required, iNaturalist can export the data in a spreadsheet for users to download. However, there is no alert that the coordinates are altered and replaced by an obscured location during this process. Although the exported spreadsheet includes a column indicating whether the location is obscured, a naïve user can easily overlook this parameter, especially when downloading a large data set that includes both threatened and nonthreatened species.

Additionally, other open-source databases (e.g., iNaturalist Network) may share data from iNaturalist, a practice we call *secondary sharing*. For example, this secondary sharing occurs every 7 days in GBIF, where the full data set from iNaturalist that is research grade and properly licensed is imported (iNaturalist, 2023). When downloading data from GBIF, it is not clearly

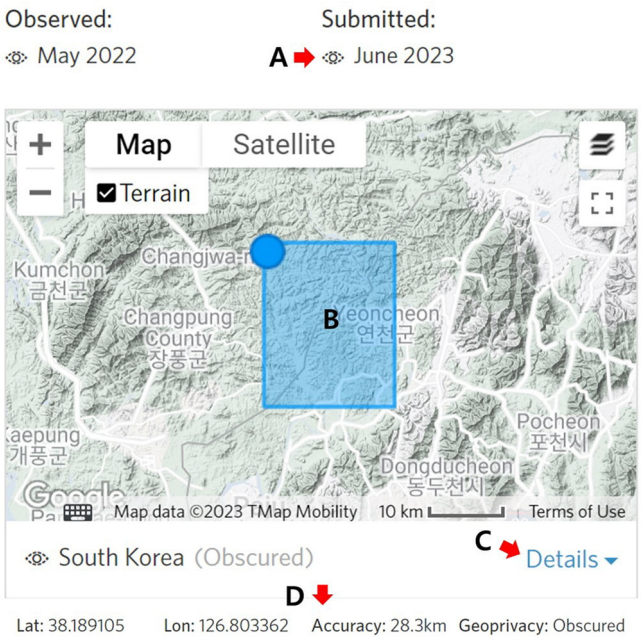


FIGURE 1 Information screen for the endangered *Mauremys reevesii* on iNaturalist (A, eye icon indicating location of the species is obscured; B, obscured location; blue box, location area within approximately 500 km^2; blue dot, point newly created by the obscured function that is not the original location of the observation; C, details tab; D, data that were obscured, including latitude, longitude, and accuracy).

disclosed that observations may contain obscured locations. Although GBIF flags potentially erroneous coordinates and provide coordinate uncertainty (in meters), it does not explicitly mark obscured data. The erroneous geographic information can lead to false positive results, which overestimate species richness in regions outside the center of the distribution range (Maldonado et al., 2015). Consequently, many publications for which the obscured data were used may have incorrect results due to the authors' lack of awareness that the coordinates were modified by the database's functions. Therefore, a comprehensive study on the impact of obscured locations on analytical results is essential to improve the use of open-source databases.

The impact of obscured data is most pronounced in studies on habitat composition and species distribution (Contreras-Díaz et al., 2023). Species distribution models (SDMs) are particularly sensitive to the geographic information of target species. Alterations in GPS locations can significantly change the predicted results for habitat type and species distribution (Rondinini et al., 2006). For example, Gábor et al. (2022) noted that the performance of SDMs decreases as errors in geographic information increase in occurrence data. To address this problem, researchers must consider positional uncertainty by implementing certain measures, such as exclusion of data points below a certain accuracy level (Finger et al., 2022) or outside the realistic distribution range (Nascimento et al., 2022; Powers & Jetz, 2019). Examples of the latter include instances where terrestrial animals are indicated to be in rivers or seas and where marine life is located on mountains. Most importantly, SDMs are commonly used in the decision-making process for

15231739, 2025, 5, Downloaded from https://conbio.onlinelibrary.wiley.com/doi/10.1111/cobi.70050 by Ewha Womans University Library, Wiley Online Library on [13/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

conservation strategies (Franklin, 2010; Villero et al., 2017). Ignoring positional uncertainty in data sets can ultimately lead to ineffective conservation actions (Loiselle et al., 2003).

When it comes to global biodiversity losses, reptiles and amphibians (herpetofauna), the most vulnerable animal group among terrestrial vertebrates, are particularly susceptible to land-cover changes and climate change due to their narrow environmental condition requirements and limited dispersal ability (Cordier et al., 2021; Dubos, Montfort, et al., 2022; Iglesias-Carrasco et al., 2023; Nori et al., 2015). The proportion of threatened species among herpetofauna is significantly higher than among birds or mammals (Cox et al., 2022; IUCN, 2023; Stuart et al., 2004). However, until recently, many species of reptiles and amphibians were not categorized or included in conservation plans due to insufficient data (Nori et al., 2015; Platts et al., 2014; Roll et al., 2017). Although open-source databases can help fill this knowledge gap, the positional inaccuracy can hinder conservation of threatened taxa (Contreras-Díaz et al., 2023). Therefore, it is essential to review whether this issue is present in literature and to raise awareness of positional inaccuracies in open-source databases within the scientific community.

We sought to determine the impact of obscured locations on research results by testing how much the obscured function can alter the original location coordinates of animals, their habitat composition, and results of species distribution modeling. To achieve this goal, we analyzed the extent to which the original data were altered using data collected from 3 endangered species in South Korea, which are characterized by high levels of habitat specialization and are severely threatened by habitat loss (Do et al., 2017, 2022). We tested whether the obscured locations could change the model predictions of their distributions and habitat types. Additionally, to understand how widespread the problem of the obscured function is in the literature, we investigated whether authors who obtained data from iNaturalist acknowledged this modification in geographic information in their articles. Although the effectiveness of the obscured function in protecting rare and endangered species is recognized, we examined potential biases when using obscured data in open-source databases.

METHODS

Study species and data collection

We selected 3 endangered species known for their specific habitat requirements or limited ranges in South Korea: the gold-spotted pond frog (*Pelophylax chosenicus*), found in lowland grasslands; Reeves' turtle (*Mauremys reevesii*), found in reservoirs and rivers; and the Mongolia racerunner (*Eremias argus*), found in coastal and river dunes (Chang et al., 2021; Do et al., 2017; Lovich et al., 2011). The location data were obtained from the National Natural Environment Survey and monitoring of invasive alien species conducted by the National Institute of Environment (NIE) under guidelines set by the Ministry of Environment (Baek et al., 2024; Do et al., 2017, 2022; NIE, 2017, 2023).

We recorded GPS coordinates and elevation (hand-held GPS device [Oregon 550; Garmin]), date, and land cover at each observation site. All data were identified and verified by us (Appendix S1). Land cover was classified (Appendix S1) as bare land (i.e., empty ground with no trees, grass, or buildings); farmland (i.e., land used for farming); forest (i.e., land covered with trees and plants); grassland (i.e., land covered with grass); urban land (i.e., city or town); water (ocean, river, reservoir, stream); and wetland (i.e., land covered with swamp or marsh).

We collected microclimate and environmental data from WorldClim 2.0 (http://www.worldclim.org; accessed on 9 August 2023) for the locations in South Korea where the species were found. All grids were set to a uniform size of 30 s (approximately 1 km^2). Following Fick and Hijmans (2017), we used 19 climatic variables derived from monthly temperature and precipitation values. Detailed descriptions of the 19 variables are at https://www.worldclim.org/data/bioclim.html, and they are listed in Table 1's footnote.

Comparisons between the original and obscured locations

We generated obscured locations by uploading our original data for the 3 species to iNaturalist. This included 65 data points for *P. chosenicus*, 75 for *M. reevesii*, and 172 for *E. argus*. The geographic information was automatically obscured by iNaturalist (i.e., field location was replaced with a random location within a 0.2 × 0.2° cell). In this manuscript, we defined the location altered by iNaturalist as the obscured location. We then downloaded the obscured location for analysis (Appendix S1). We used the following measures to examine the differences between the original and obscured data: linear distance between the field location and the obscured location in Google Earth, elevation, land cover, changes in habitat microclimate and environmental variables, and predicted species distribution. We tested the differences in elevation between the original and obscured data with a paired *t* test. The Shapiro test was employed to assess the assumption of normality. When the assumption was not met, we used paired Wilcoxon signed-rank test to examine the differences. The Levene test was used to examine the variance between the original and obscured data. Pearson's chi-square test was used to examine differences in land-cover proportions between the original and obscured data.

We conducted species distribution modeling with Maxent 3.4.4, a machine learning-based modeling software (Phillips & Dudík, 2008). To avoid multicollinearity issues, we excluded correlated variables with Pearson's correlation coefficients (r) >0.7 (Dormann et al., 2013). Spatial thinning was not performed because it can degrade the performance of SDMs in rare species with limited data (Lamboley & Fourcade, 2024; Steen et al., 2021; Ten Caten & Dallas, 2023). Each model was run 10 times with default parameters and logistic output and 10,000 background points (Do et al., 2022). To evaluate the models, we randomly divided the data set into a training set (75%) and a testing set (25%) with the random test percentage feature. We

15231739, 2025, 5, Downloaded from https://conbio.onlinelibrary.wiley.com/doi/10.1111/cobi.70050 by Ewha Womans University Library, Wiley Online Library on [13/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Difference in the area under the curve (AUC) and the percentage contributions of climatic variables to species distribution models between filed data and obscured data.

Bioclimatic variables[a]	*Pelophylax chosenicus*		*Mauremys reevesii*		*Eremias argus*	
	Original	Obscured	Original	Obscured	Original	Obscured
(AUC)	0.959	0.914	0.719	0.953	0.925	0.990
Bio01	29.9	31.5	63.4	3.8	2.3	15.3
Bio02	10	10	4.6	2.6	52.5	50.8
Bio03				10.1	26.7	
Bio04	48.7	44.9				
Bio09				30.0		
Bio10						
Bio11					4.7	
Bio12	11.4	13.6	7.8	17.0		
Bio13			0.2			
Bio14			23.9	35.8		9.7
Bio15						9.8
Bio17					13.8	
Bio18						
Bio19						8.5

[a]Bioclimatic variables: Bio01, annual mean temperature; Bio2, mean diurnal range (mean of monthly [max temp − min temp]); Bio3, isothermality (Bio2/Bio7) (×100); Bio4, temperature seasonality (SD × 100); Bio5, max temperature of warmest month; Bio6, min temperature of coldest month; Bio7, temperature annual range (Bio5–Bio6); Bio8, mean temperature of wettest quarter; Bio9, mean temperature of driest quarter; Bio10, mean temperature of warmest quarter; Bio11, mean temperature of coldest quarter; Bio12, annual precipitation; Bio13, precipitation of wettest month; Bio14, precipitation of driest month; Bio15, precipitation seasonality (coefficient of variation); Bio16, precipitation of wettest quarter; Bio17, precipitation of driest quarter; Bio18, precipitation of warmest quarter; Bio19, precipitation of coldest quarter.

displayed the contributing variables and the area under the curve (AUC) value, which indicates the model's explanatory power, to illustrate the differences in modeling results between the original and obscured data (Swets, 1988).

Literature search and assessment

We searched for papers published from January 2014, the earliest appearance of iNaturalist in the literature, to December 2022 to investigate whether authors acknowledged that their data had been altered by iNaturalist's obscured function (Appendix S2). We searched Web of Science and Google Scholar for keywords *iNaturalist* and *modeling*. For a paper to be relevant to our study, authors had to have directly downloaded data from iNaturalist for analyses. We included only peer-reviewed papers in English. Papers were excluded if the target species was not identified or if the source of the acquired data was not clearly specified.

Estimating awareness of obscured locations

For the collected papers, we verified whether each paper actually used data from iNaturalist. Then, we examined the geoprivacy status (open, obscured, or not specified) of the studied species in all papers. If the geoprivacy was obscured, we checked whether the authors acknowledged the obscured function and whether they requested the original data from the uploader. In iNaturalist, species classified as rare or endangered are automatically obscured by the database. Therefore, if the species was not endangered, we checked whether the uploader had manually obscured the geographic information. Our literature assessment may have included some studies by researchers who were aware of the obscured function but did not disclose whether they had excluded the obscured data.

RESULTS

Impact of obscured data

The obscured locations differed substantially from the original locations for all 3 endangered species. The average distance between the original and obscured locations was 10.3 km (SD 5.0) ($n = 65$) for *P. chosenicus*, 13.0 km (5.9) ($n = 75$) for *M. reevesii*, and 12.5 km (5.3) ($n = 172$) for *E. argus*. We found significant differences in elevation between the original and obscured locations for all 3 species (*P. chosenicus* effect size = 28.31; *M. reevesii* effect size = −0.74; *E. argus* effect size = 7.07) (Figure 2). Compared with the original locations, the elevation of *P. chosenicus* and *M. reevesii* significantly increased, whereas the elevation of *E. argus* significantly decreased in the obscured locations (Figure 2). Furthermore, the elevation for all 3 species showed significantly greater variability with the obscured data than with original data

15231739, 2025, 5, Downloaded from https://conbio.onlinelibrary.wiley.com/doi/10.1111/cobi.70050 by Ewha Womans University Library, Wiley Online Library on [13/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

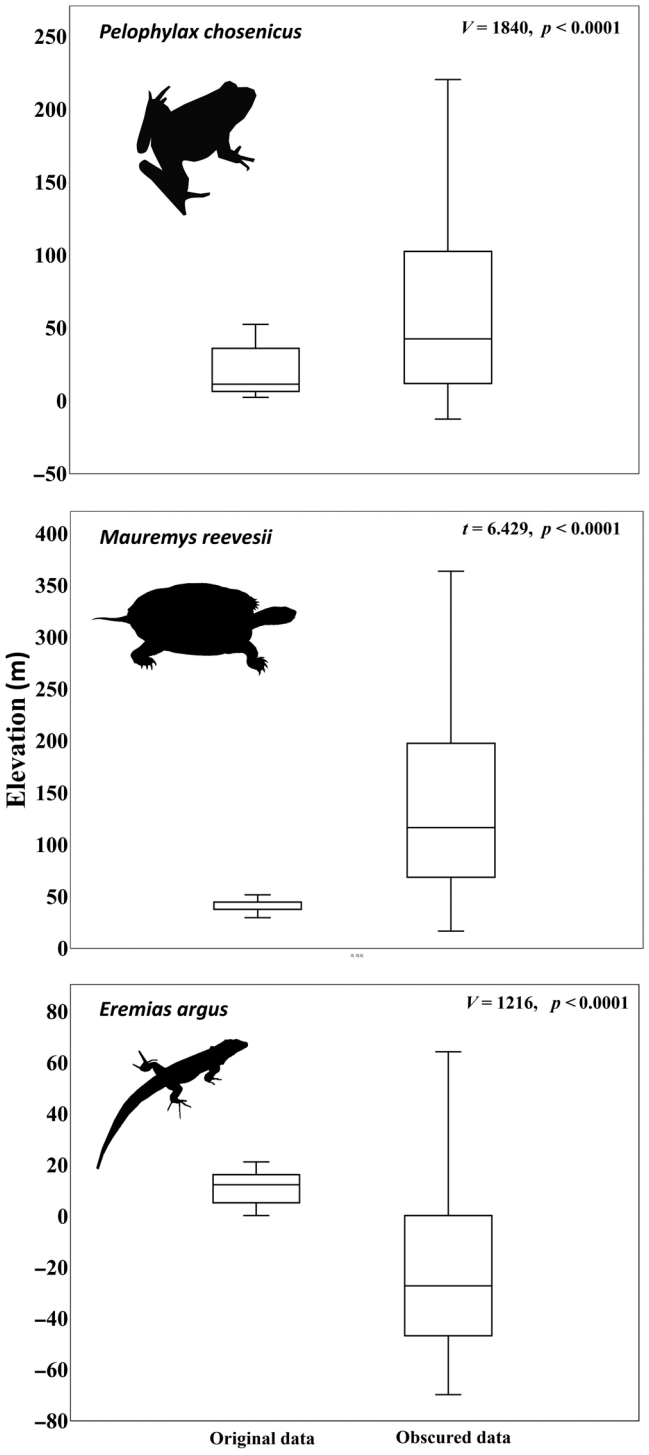


FIGURE 2 Elevation distributions based on original data (original data) and on obscured locations (original coordinates replaced with a random location in a 0.2 × 0.2° cell) from iNaturalist in 3 endangered species (horizontal line, median; box ends, interquartile range [IQR]; whiskers, range of data [or maximum and minimum]). The *V* is derived from the paired Wilcoxon signed rank test, and the *t* is derived from paired *t* test).

(*P. chosenicus* $F = 23.67$, df = 1, $p < 0.0001$, effect size = −1.64; *M. reevesii* $F = 12.56$, df = 1, $p < 0.0001$, effect size = −0.05; *E. argus* $F = 200.46$, df = 1, $p < 0.0001$, effect size = −1.63). With the obscured data, the location for *E. argus* extended to areas under water and beyond the natural geographic distribution of the species.

The 3 most preferred types of land cover for the gold-spotted pond frog (*P. chosenicus*) were farmland (57%), wetland (29%), and water (5%) based on original data, whereas they were forest (31%), farmland (20%), and urban (20%) based on obscured data (Figure 3). *Pelophylax chosenicus* did not occur in forest in the original data. For the Reeves' turtle (*M. reevesii*), the preferred land covers were water (63%) and urban (24%) based on original data and forest (71%) and farmland (15%) based on obscured data. Similarly, the Mongolia racerunner (*E. argus*) was primarily found on bare land (78%) based on original data, but the most dominant land cover was water (80%) in the obscured data. The land cover of habitats was significantly different between the original and obscured data for all 3 species (Pearson's chi-square test $p < 0.001$ in all cases, effect size 0.71, 0.89, and 0.91 for *P. chosenicus*, *M. reevesii*, and *E. argus*, respectively).

The environmental variables contributing to the SDMs differed substantially between the original and obscured data (Table 1) in *M. reevesii* and *E. argus*. For *M. reevesii*, the main contributor changed from annual mean temperature (Bio01) to precipitation of driest month (Bio14). In the case of *E. argus*, the top contributor was mean diurnal range (Bio02), but the following 2 contributors changed from isothermality (Bio03) and precipitation of driest quarter (Bio17) to annual mean temperature (Bio01) and precipitation seasonality (Bio15). We did not find any differences between the original and obscured data in *P. chosenicus* relative to the 3 factors that contributed the most.

The AUC values differed between models based on original data and models based on obscured data (Table 1). For *P. chosenicus*, the model with obscured data had showed a lower AUC value (0.914) than the model with original data (0.959). In contrast, for the other 2 species, the AUC values were greater with obscured data (*M. reevesii* 0.953, *E. argus* 0.990) than with original data (*M. reevesii* 0.719, *E. argus* 0.925).

The potential distributions of the 3 species were also markedly different between the original and obscured data (Figure 4). For *P. chosenicus*, the potential distribution based on obscured data was much broader than that based on the original data. The potential distribution of *M. reevesii* based on obscured data was mostly restricted to the southern corner of the nation, whereas the potential distribution based on original data extended across the country, except for high mountains. Finally, *E. argus* was found mostly in the northern and southern parts of the country according to the original data. However, based on the obscured data, its distribution was confined to the tip of one peninsula. Thus, species distributions were vastly misleading and, when based only on obscured data, extended into areas where the species are not typically found.

15231739, 2025, 5, Downloaded from https://conbio.onlinelibrary.wiley.com/doi/10.1111/cobi.70050 by Ewha Womans University Library, Wiley Online Library on [13/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

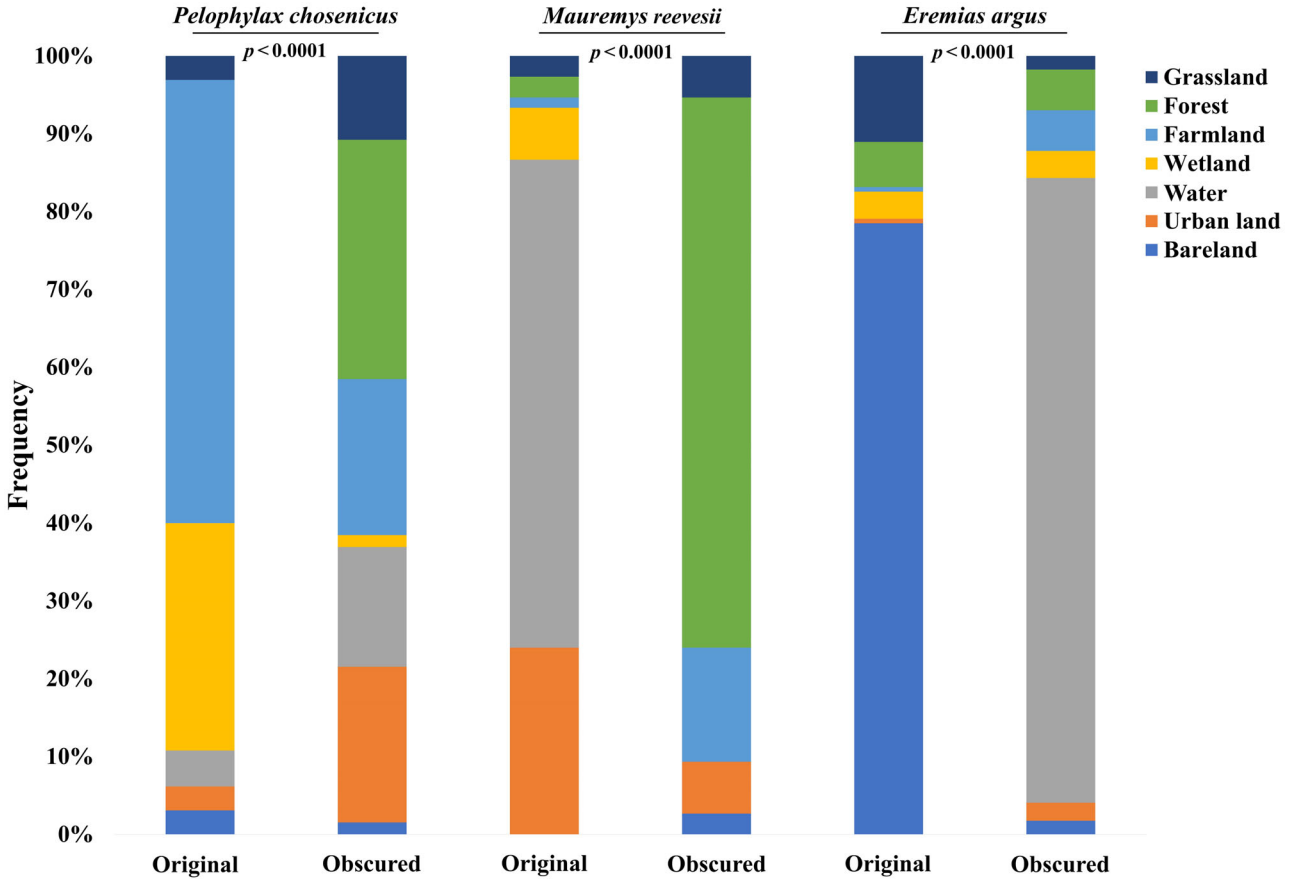


FIGURE 3 Comparison of land-cover types from original data (original) and obscured (land cover at a random location in a 0.2 × 0.2° cell) data for 3 endangered species. Obscured data were generated by entering original data under the iNaturalist obscured function (*p* values based on Pearson's chi-square test).

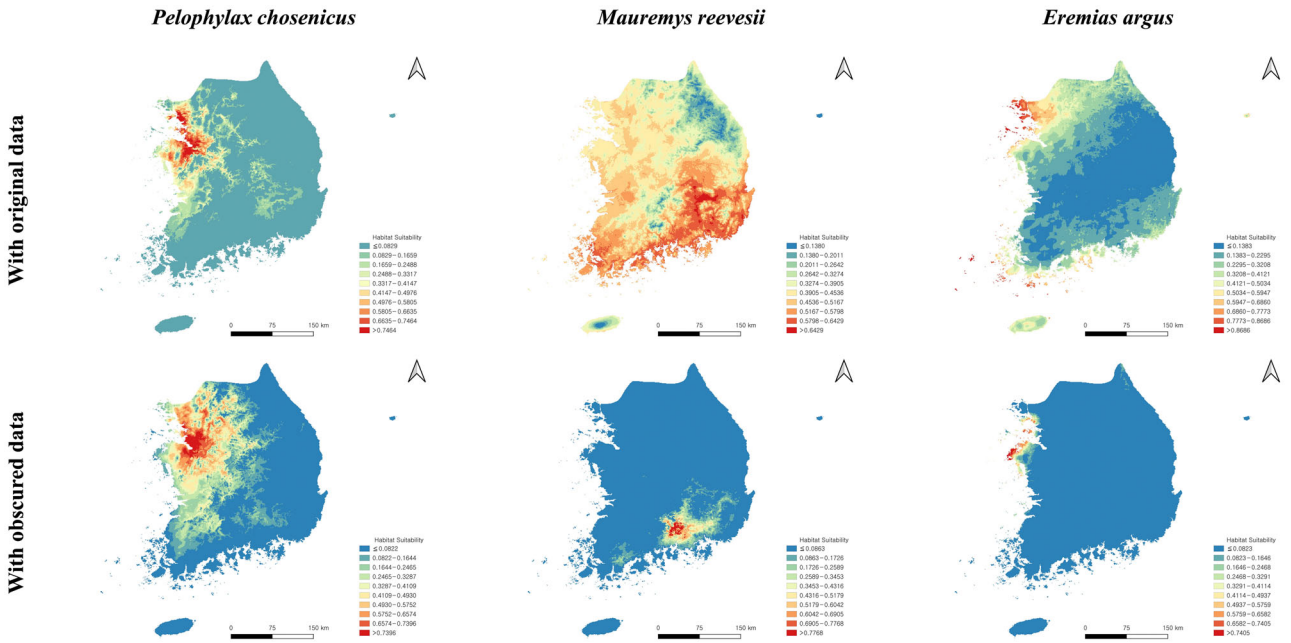


FIGURE 4 Potential distribution of 3 endangered Korean species based on original data and on iNaturalist obscured data (original coordinates replaced with a random location in a 0.2 × 0.2° cell) (the redder the area, the higher the likelihood of the species being found).

15231739, 2025, 5, Downloaded from https://conbio.onlinelibrary.wiley.com/doi/10.1111/cobi.70050 by Ewha Womans University Library, Wiley Online Library on [13/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Lack of awareness of use of obscured data

A total of 705 papers met the criteria for our literature search, and for 228 papers (32.3%), researchers used iNaturalist data (Appendix S2). Among the 228 papers, 74.7% ($n = 170$) of the modeling studies included obscured data, and the rest only referred to pictures and locations from iNaturalist without modeling. Of these 170 papers, only 4 studies (2.4%) acknowledged the existence of the obscured function and excluded obscured data from analyses. Eight studies did not recognize the obscured function but excluded certain GPS locations that appeared in freshwater ($n = 1$), on islands ($n = 1$), at sea ($n = 5$), and in coastal areas ($n = 1$), which were presumably outside the typical distribution of the species in question. None of the studies using obscured data requested the original data from the uploader.

DISCUSSION

Obscured function protection of threatened species

The obscured function in iNaturalist significantly altered the locations of specimens, effectively moving them away from their original locations. In this sense, the obscured function partially achieved its goal of concealing the original locations of endangered species. However, the information derived from the obscured locations, such as land-cover type, microclimate, and species distribution, was misleading and did not reflect the true conditions of the endangered species we examined.

Significant changes in habitat information

Habitat types based on obscured data are likely to be erroneous for species found in limited areas or those with specific environmental requirements. For example, the golden-spotted pond frogs are typically found in wetlands and, in South Korea, are primarily restricted to rice paddies (Do et al., 2022; Oh et al., 2021). Therefore, the original data correctly identified their habitat types as farmland and wetland. However, the obscured data incorrectly identified their main habitat type as forest, where these frogs do not occur at all. Due to their specific habitat requirements, the habitat types based on obscured data were inaccurate.

The situation with *M. reevesii* and *E. argus* was even more problematic. *Mauremys reevesii* lives in water systems, such as reservoirs, rivers, and channels (Do et al., 2022; Lin et al., 2015). The largest population of *M. reevesii* dwells in Geumhoji Pond, which is only 0.2 km^2 and is in Jinju province (Kang et al., 2025). The rivers where this species is found are typically a few hundred meters wide. Thus, the main habitat type for *M. reevesii*, which should be water, was incorrectly designated as forest in the obscured data. Reservoirs and rivers in Korea are usually adjacent to forests. Similarly, the Mongolia racerunner occurs only in sandy areas (bare land) (Chang et al., 2021; Do et al., 2022). In South Korea, these sandy areas are narrowly distributed along water bodies, such as seas or rivers. When the obscured function was applied, some data points were placed in the sea, where the Mongolia racerunner does not occur at all. Furthermore, the main habitat type for the Mongolia racerunner changed from sandy areas with original data to water with obscured data.

Both Reeves' turtle and Mongolian racerunner occur in specific and limited habitat types, and the obscured data did not reflect their true habitats. This problem has been noted elsewhere, such as when the distribution of the Andean bumble bee (*Bombus funebris*) was placed in the ocean by the obscured function (Nascimento et al., 2022). In these cases, the error could be recognized because the points were in unlikely landscapes where the species do not occur (Chapman et al., 2019; Goodman et al., 2022; Tran et al., 2022). However, in cases where the error cannot be clearly seen, thus impeding modification or explanation, the potential impact arising from obscured locations may become even greater.

The obscured function can affect the AUC value of SDMs. The AUC value varies depending on the geographic extent of the study area. As the study area expands beyond the actual distribution range of the species, the AUC value tends to be overestimated (Acevedo et al., 2012; Leroy et al., 2018; Lobo et al., 2008; Luoto et al., 2005). This overestimation occurs because the model includes accurate predictions of absence in areas unrelated to the actual distribution of the species. The model's results and explanatory power (as measured by AUC) can change significantly when the obscured function alters the apparent distribution of the species. Moreover, AUC values can exaggerate model performance for rare species, especially in cases of imbalanced data (e.g., when presence data are very limited) (Breiner et al., 2015; Jiménez-Valverde et al., 2009; Vaughan & Ormerod, 2005). In such situations, AUC values may not accurately reflect the actual performance of the model (Dubos, Préau, et al., 2022). For endangered species with few coordinate points, applying the obscured function may lead to even more dramatic changes in model results. This is because the limited data points for rare species make the models more sensitive to spatial alterations introduced by the obscuring process as shown in our study.

Possibility of not recognizing the obscured function

Our literature assessment revealed that many authors may not have accounted for the inaccuracies introduced by the obscured function when using data from iNaturalist. Only a small fraction of studies avoided the pitfalls of directly downloading obscured data from iNaturalist for their analyses. The problem is even more pronounced in cases of secondary data sharing, where researchers obtained data from databases that use iNaturalist data. For example, from 2011 to the present, most of the GBIF available occurrences of vertebrates listed as threatened on the International Union for Conservation of Nature Red List are provided by iNaturalist (Contreras-Díaz et al., 2023). Given that the exported datasheet does not explicitly specify whether the

15231739, 2025, 5, Downloaded from https://conbio.onlinelibrary.wiley.com/doi/10.1111/cobi.70050 by Ewha Womans University Library, Wiley Online Library on [13/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

obscured function is applied, users need to pay extra attention to the coordinate uncertainties in the exported datasheet. Ideally, authors should apply data cleaning procedures, such as using CoordinateCleaner or bdc packages in R (Ribeiro et al., 2022; Zizka et al., 2019), before running analyses. This might mitigate the impacts of obscured locations if they are removed during these steps.

Because we restricted awareness of use of obscured data to papers that directly imported their data from iNaturalist, potentially thousands of papers may have used secondarily shared data from GBIF. Contreras-Díaz et al. (2023) demonstrated that a change in even 25% of the data can cause significant misestimation of geographic distribution and ecological niche. This misestimation can occur in both threatened and nonthreatened species; errors are likely higher for studies involving rare and threatened species, where data are automatically obscured by iNaturalist. Using these erroneous spatial data in biodiversity studies could lead to incorrect decisions on conservation priorities for wildlife (Araújo et al., 2019; Trisurat et al., 2012). Although numerous researchers may have been aware of the obscured function and some may have even addressed it, there remains a significant possibility that some researchers used obscured data without fully understanding its implications. Due to this potential oversight, it is imperative to conduct a comprehensive reexamination of the processes governing data reproduction and exportation in open-source databases. Such a review is crucial to ensure the integrity and reliability of research outcomes derived from these widely used resources.

Solutions to the problem

We propose several actions to prevent errors associated with using data from iNaturalist. First, there is a need to actively promote and increase awareness of the obscured function. Although iNaturalist's help section clearly explains this function, users may still overlook it. Therefore, iNaturalist should issue warnings to all users to treat the data with caution. One approach could be to display a pop-up window when users download obscured data, indicating that the coordinates have been altered and requiring confirmation of understanding from the downloader. Alternatively, GBIF (and other data repositories) should specifically label coordinates as potentially erroneous or imprecise in the exported data set. Another method could involve organizing a group of experts to review requests for raw data, similar to iNaturalist's current data grading system, and then grant access to the database accordingly. For this to be feasible, the uploader or original author would need to grant iNaturalist permission to freely use and license the material.

In conclusion, our results empirically demonstrate that using obscured data in species distribution modeling can lead to erroneous outcomes. Moreover, the result of SDMs could directly influence the determination of conservation targets, scope, and priorities. Our research underscores the necessity for individual researchers and the scientific community to thoroughly vet and examine data obtained from open-source databases. Researchers must carefully read the database guidelines, and such data should be accompanied by sufficient explanations to prevent potential errors in subsequent analyses. However, much conservation and ecology data are fluid. Therefore, it is plausible that our results could be subject to modification should the conservation status of the species in question shift in the future. If a common species becomes a protected species, its coordinate values may subsequently be affected by the obscured function. This should be a consideration by all researchers whether they are using citizen science data for research or reading reviews based on said data.

ACKNOWLEDGMENTS

This work was supported by a grant from the Rural Administration of Korea, funded by the Korean government (MAFRA) (RS-2024-00397542) and Korea Environmental Industry & Technology Institute (KEITI)(grant number:RS-2021-KE001530, RS-2021-KE001362).

ORCID

Kyo Soung Koo https://orcid.org/0000-0003-0294-0875
Ko-Huan Lee https://orcid.org/0000-0001-5577-8619
Dawon Lee https://orcid.org/0009-0009-1749-2949
Yikweon Jang https://orcid.org/0000-0002-7213-7319

REFERENCES

Acevedo, P., Jiménez-Valverde, A., Lobo, J. M., & Real, R. (2012). Delimiting the geographical background in species distribution modelling. *Journal of Biogeography*, *39*(8), 1383–1390.

Araújo, M. B., Anderson, R. P., Márcia Barbosa, A., Beale, C. M., Dormann, C. F., Early, R., Garcia, R. A., Guisan, A., Maiorano, L., Naimi, B., O'Hara, R. B., Zimmermann, N. E., & Rahbek, C. (2019). Standards for distribution models in biodiversity assessments. *Science Advances*, *5*(1), Article eaat4858.

Baek, H. J., Cheong, E., Kim, Y., Koo, K. S., Kim, S. H., Park, C. D., & Yoon, J. D. (2024). Geographical distribution of *Mauremys sinensis*, *Mauremys reevesii*, and their hybrids in South Korea. *Animals*, *14*(18), Article 2626.

Ball-Damerow, J. E., Brenskelle, L., Barve, N., Soltis, P. S., Sierwald, P., Bieler, R., Lafrance, R., Ariño, A. H., & Guralnick, R. P. (2019). Research applications of primary biodiversity databases in the digital age. *PLoS ONE*, *14*(9), Article e0215794.

Bisby, F. A. (2000). The quiet revolution: Biodiversity informatics and the internet. *Science*, *289*(5488), 2309–2312.

Breiner, F. T., Guisan, A., Bergamini, A., & Nobis, M. P. (2015). Overcoming limitations of modelling rare species by using ensembles of small models. *Methods in Ecology and Evolution*, *6*(10), 1210–1218.

Campbell, C. J., Barve, V., Belitz, M. W., Doby, J. R., White, E., Seltzer, C., Di Cecco, G., Hurlbert, A. H., & Guralnick, R. (2023). Identifying the identifiers: How iNaturalist facilitates collaborative, research-relevant data generation and why it matters for biodiversity science. *Bioscience*, *73*(7), 533–541.

Chang, M. H., Song, J. Y., & Koo, K. S. (2021). Effect of coastal dune restoration on the population of endangered Mongolian racerunner (*Eremias argus*) in the Republic of Korea. *Journal of Coastal Conservation*, *25*(2), Article 29.

Chapman, D., Pescott, O. L., Roy, H. E., & Tanner, R. (2019). Improving species distribution models for invasive non-native species with biologically informed pseudo-absence selection. *Journal of Biogeography*, *46*(5), 1029–1040.

Contreras-Díaz, R. G., Nori, J., Chiappa-Carrara, X., Peterson, A. T., Soberón, J., & Osorio-Olvera, L. (2023). Well-intentioned initiatives hinder understanding biodiversity conservation: Cloaked iNaturalist information for threatened species. *Biological Conservation*, *282*, Article 110042.

15231739, 2025, 5, Downloaded from https://conbio.onlinelibrary.wiley.com/doi/10.1111/cobi.70050 by Ewha Womans University Library, Wiley Online Library on [13/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Cordier, J. M., Aguilar, R., Lescano, J. N., Leynaud, G. C., Bonino, A., Miloch, D., Loyola, R., & Nori, J. (2021). A global assessment of amphibian and reptile responses to land-use changes. *Biological Conservation*, *253*, Article 108863.

Costello, M. J., Appeltans, W., Bailly, N., Berendsohn, W. G., De Jong, Y., Edwards, M., Froese, R., Huettmann, F., Los, W., Mees, J., Segers, H., & Bisby, F. A. (2014). Strategies for the sustainability of online open-access biodiversity databases. *Biological Conservation*, *173*, 155–165.

Costello, M. J., Michener, W. K., Gahegan, M., Zhang, Z. Q., & Bourne, P. E. (2013). Biodiversity data should be published, cited, and peer reviewed. *Trends in Ecology & Evolution*, *28*(8), 454–461.

Cox, N., Young, B. E., Bowles, P., Fernandez, M., Marin, J., Rapacciuolo, G., Böhm, M., Brooks, T. M., Hedges, S. B., Hilton-Taylor, C., Hoffmann, M., Jenkins, R. K. B., Tognelli, M. F., Alexander, G. J., Allison, A., Ananjeva, N. B., Auliya, M., Avila, L. J., Chapple, D. G., … Xie, Y. (2022). A global reptile assessment highlights shared conservation needs of tetrapods. *Nature*, *605*, 285–290.

Do, M. S., Lee, J. W., Jang, H. J., Kim, D. I., Park, J., & Yoo, J. C. (2017). Spatial distribution patterns and prediction of hotspot area for endangered herpetofauna species in Korea. *Korean Journal of Environment and Ecology*, *31*(4), 381–396.

Do, M. S., Son, S. J., Choi, G., Yoo, N., Kim, D. I., Koo, K. S., & Nam, H. K. (2022). The establishment of ecological conservation for herpetofauna species in hotspot areas of South Korea. *Scientific Reports*, *12*(1), Article 14839.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., Mcclean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27–46.

Dubos, N., Montfort, F., Grinand, C., Nourtier, M., Deso, G., Probst, J.-M., Razafimanahaka, J. H., Andriantsimanarilafy, R. R., Rakotondrasoa, E. F., Razafindraibe, P., Jenkins, R., & Crottini, A. (2022). Are narrow-ranging species doomed to extinction? Projected dramatic decline in future climate suitability of two highly threatened species. *Perspectives in Ecology and Conservation*, *20*(1), 18–28.

Dubos, N., Préau, C., Lenormand, M., Papuga, G., Monsarrat, S., Denelle, P., Louarn, M. L., Heremans, S., May, R., Roche, P., & Luque, S. (2022). Assessing the effect of sample bias correction in species distribution models. *Ecological Indicators*, *145*, Article 109487.

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315.

Finger, N., Farleigh, K., Bracken, J. T., Leaché, A. D., François, O., Yang, Z., Flouri, T., Charran, T., Jezkova, T., Williams, D. A., & Blair, C. (2022). Genome-scale data reveal deep lineage divergence and a complex demographic history in the Texas horned lizard (*Phrynosoma cornutum*) throughout the southwestern and central United States. *Genome Biology and Evolution*, *14*(1), Article evab260.

Fisher, S., Fisher, R. N., & Pauly, G. B. (2022). Hidden in plain sight: Detecting invasive species when they are morphologically similar to native species. *Frontiers in Conservation Science*, *3*, Article 846431.

Franklin, J. (2010). *Mapping species distributions: Spatial inference and prediction*. Cambridge University Press.

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*(10), 2289–2302.

Goodman, A., Allen, J., Brim, J., Codella, A., Hahn, B., Jojo, H., Bondocgawa Mafla-Mills, Z., Bondoc Mafla, S. T., Oduro, A., Wilson, M., & Ware, J. (2022). Utilization of community science data to explore habitat suitability of basal termite genera. *Insect Systematics and Diversity*, *6*(4), Article 5.

Hochkirch, A., Samways, M. J., Gerlach, J., Böhm, M., Williams, P., Cardoso, P., Cumberlidge, N., Stephenson, P. J., Seddon, M. B., Clausnitzer, V., Borges, P. A. V., Mueller, G. M., Pearce-Kelly, P., Raimondo, D. C., Danielczak, A., & Dijkstra, K.-D. B. (2021). A strategy for the next decade to address data deficiency in neglected biodiversity. *Conservation Biology*, *35*(2), 502–509.

Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven shortfalls that beset large-scale knowledge of biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, *46*(1), 523–549.

Iglesias-Carrasco, M., Medina, I., & Ord, T. J. (2023). Global effects of forest modification on herpetofauna communities. *Conservation Biology*, *37*(1), Article e13998.

iNaturalist. (2023). *Homepage*. https://www.inaturalist.org

IUCN. (2023). *Homepage*. https://www.iucnredlist.org

Jiménez-Valverde, A., Lobo, J., & 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.

Kang, H., Oh, J. M., Jang, Y., & Koo, K. S. (2025). Population size and structure of endangered Reeves' turtle (*Mauremys reevesii*) in Geumho Reservoir, South Korea: Possibility of the largest population. *Korean Journal of Environmental Biology*, *43*(1), 1–9.

Koldasbayeva, D., Tregubova, P., Shadrin, D., Gasanov, M., & Pukalchik, M. (2022). Large-scale forecasting of *Heracleum sosnowskyi* habitat suitability under the climate change on publicly available data. *Scientific Reports*, *12*(1), Article 6128.

Lamboley, Q., & Fourcade, Y. (2024). No optimal spatial filtering distance for mitigating sampling bias in ecological niche models. *Journal of Biogeography*, *51*(9), 1783–1794.

Lane, M. A., & Edwards, J. L. (2007). The global biodiversity information facility (GBIF). *Systematics Association Special Volume*, *73*, Article 1.

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.

Lin, S. M., Lee, Y., Chen, T. H., & Lin, J. W. (2015). Habitat preference and management of a Chinese pond turtle population protected by the demilitarized Kinmen Islands. *Journal of Herpetology*, *49*(3), 399–404.

Lindenmayer, D., & Scheele, B. (2017). Do not publish. *Science*, *356*(6340), 800–801.

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.

Loiselle, B. A., Howell, C. A., Graham, C. H., Goerck, J. M., Brooks, T., Smith, K. G., & Williams, P. H. (2003). Avoiding pitfalls of using species distribution models in conservation planning. *Conservation Biology*, *17*(6), 1591–1600.

Lovich, J., Yasukawa, Y., & Ota, H. (2011). *Mauremys reevesii* (Gray 1831)–Reeves' turtle, Chinese three-keeled pond turtle. In A. G. J. Rhodin, P. C. H. Pritchard, P. P. van Dijk, R. A. Saumure, K. A. Buhlmann, J. B. Iverson, & R. A. Mittermeier (Eds.), *Conservation biology of freshwater turtles and tortoises* (Vol. 5, pp. 050.1–050.10). Chelonian Research Foundation.

Luoto, M., Pöyry, J., Heikkinen, R. K., & Saarinen, K. (2005). Uncertainty of bioclimate envelope models based on the geographical distribution of species. *Global Ecology and biogeography*, *14*(6), 575–584.

Maldonado, C., Molina, C. I., Zizka, A., Persson, C., Taylor, C. M., Albán, J., Chilquillo, E., Rønsted, N., & Antonelli, A. (2015). Estimating species diversity and distribution in the era of Big Data: To what extent can we trust public databases? *Global Ecology and Biogeography*, *24*(8), 973–984.

Nascimento, A. C., Montalva, J., Ascher, J. S., Engel, M. S., & Silva, D. P. (2022). Current and future distributions of a native Andean bumble bee. *Journal of Insect Conservation*, *26*(4), 559–569.

National Institute of Ecology (NIE). (2017). *30 years of the Natural Environment Survey: 1986–2015*. Seocheon, Designglory.

National Institute of Ecology (NIE). (2023). *Guidelines for the 6th National Natural Environment Survey*. Seocheon, Designglory.

Nori, J., Lemes, P., Urbina-Cardona, N., Baldo, D., Lescano, J., & Loyola, R. (2015). Amphibian conservation, land-use changes and protected areas: A global overview. *Biological Conservation*, *191*, 367–374.

Nugent, J. (2018). iNaturalist: Citizen science for 21st-century naturalists. *Science Scope*, *41*(7), 12–15.

Oh, H.-J., Chang, K.-H., Jin, M.-Y., Suh, J.-M., Yoon, J.-D., Shin, K.-H., Park, S.-G., & Chang, M.-H. (2021). Trophic ecology of endangered gold-spotted pond frog in ecological wetland park and rice paddy habitats. *Animals*, *11*(4), Article 967.

15231739, 2025, 5, Downloaded from https://conbio.onlinelibrary.wiley.com/doi/10.1111/cobi.70050 by Ewha Womans University Library, Wiley Online Library on [13/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Phillips, S. J., & Dudík, M. (2008). Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. *Ecography*, *31*(2), 161–175.

Pimm, S. L., Alibhai, S., Bergl, R., Dehgan, A., Giri, C., Jewell, Z., Joppa, L., Kays, R., & Loarie, S. (2015). Emerging technologies to conserve biodiversity. *Trends in Ecology & Evolution*, *30*(11), 685–696.

Platts, P. J., Garcia, R. A., Hof, C., Foden, W., Hansen, L. A., Rahbek, C., & Burgess, N. D. (2014). Conservation implications of omitting narrow-ranging taxa from species distribution models, now and in the future. *Diversity and Distributions*, *20*(11), 1307–1320.

Poisson, A. C., McCullough, I. M., Cheruvelil, K. S., Elliott, K. C., Latimore, J. A., & Soranno, P. A. (2020). Quantifying the contribution of citizen science to broad-scale ecological databases. *Frontiers in Ecology and the Environment*, *18*(1), 19–26.

Powers, R. P., & Jetz, W. (2019). Global habitat loss and extinction risk of terrestrial vertebrates under future land-use-change scenarios. *Nature Climate Change*, *9*(4), 323–329.

Ribeiro, B. R., Velazco, S. J. E., Guidoni-Martins, K., Tessarolo, G., Jardim, L., Bachman, S. P., & Loyola, R. (2022). bdc: A toolkit for standardizing, integrating and cleaning biodiversity data. *Methods in Ecology and Evolution*, *13*(7), 1421–1428.

Roll, U., Feldman, A., Novosolov, M., Allison, A., Bauer, A. M., Bernard, R., Böhm, M., Castro-Herrera, F., Chirio, L., Collen, B., Colli, G. R., Dabool, L., Das, I., Doan, T. M., Grismer, L. L., Hoogmoed, M., Itescu, Y., Kraus, F., LeBreton, M., … Meiri, S. (2017). The global distribution of tetrapods reveals a need for targeted reptile conservation. *Nature Ecology & Evolution*, *1*(11), 1677–1682.

Rondinini, C., Wilson, K. A., Boitani, L., Grantham, H., & Possingham, H. P. (2006). Tradeoffs of different types of species occurrence data for use in systematic conservation planning. *Ecology Letters*, *9*(10), 1136–1145.

Rowley, J. J., & Callaghan, C. T. (2020). The FrogID dataset: Expert-validated occurrence records of Australia's frogs collected by citizen scientists. *ZooKeys*, *912*, 139–151.

Shah, N., Nute, M. G., Warnow, T., & Pop, M. (2019). Misunderstood parameter of NCBI BLAST impacts the correctness of bioinformatics workflows. *Bioinformatics*, *35*(9), 1613–1614.

Steen, V. A., Tingley, M. W., Paton, P. W., & Elphick, C. S. (2021). Spatial thinning and class balancing: Key choices lead to variation in the performance of species distribution models with citizen science data. *Methods in Ecology and Evolution*, *12*(2), 216–226.

Stuart, S. N., Chanson, J. S., Cox, N. A., Young, B. E., Rodrigues, A. S., Fischman, D. L., & Waller, R. W. (2004). Status and trends of amphibian declines and extinctions worldwide. *Science*, *306*, 1783–1786.

Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D., & Kelling, S. (2009). eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, *142*(10), 2282–2292.

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, *240*(4857), 1285–1293.

Tang, B., Clark, J. S., & Gelfand, A. E. (2021). Modeling spatially biased citizen science effort through the eBird database. *Environmental and Ecological Statistics*, *28*(3), 609–630.

Ten Caten, C., & Dallas, T. (2023). Thinning occurrence points does not improve species distribution model performance. *Ecosphere*, *14*(12), Article e4703.

Tran, T. T., Carter, B. E., & Castillo Vardaro, J. A. (2022). Predicted threats to a native squirrel from two invading species based on citizen science data. *Biological Invasions*, *24*(11), 3539–3553.

Trisurat, Y., Bhumpakphan, N., Reed, D. H., & Kanchanasaka, B. (2012). Using species distribution modeling to set management priorities for mammals in northern Thailand. *Journal for Nature Conservation*, *20*(5), 264–273.

Vaughan, I. P., & Ormerod, S. J. (2005). The continuing challenges of testing species distribution models. *Journal of Applied Ecology*, *42*(4), 720–730.

Villero, D., Pla, M., Camps, D., Ruiz-Olmo, J., & Brotons, L. (2017). Integrating species distribution modelling into decision-making to inform conservation actions. *Biodiversity and Conservation*, *26*, 251–271.

Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Duarte Ritter, C., Edler, D., Farooq, H., Herdean, A., Ariza, M., Scharn, R., Svantesson, S., Wengström, 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.

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: Koo, K. S., Lee, K.-H., Lee, D., & Jang, Y. (2025). Impact of obscured data on species distribution models. *Conservation Biology*, *39*, e70050. https://doi.org/10.1111/cobi.70050



15231739, 2025, 5, Downloaded from https://conbio.onlinelibrary.wiley.com/doi/10.1111/cobi.70050 by Ewha Womans University Library, Wiley Online Library on [13/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License