

ORIGINAL RESEARCH

Spatial and taxonomic biases in bat records: Drivers and conservation implications in a megadiverse country

Veronica Zamora-Gutierrez^{1,2,3}  | Tatsuya Amano^{2,4,5}  | Kate E. Jones^{3,6} 

¹CONACYT – Centro Interdisciplinario de Investigación para el Desarrollo Integral Regional (CIIDIR) Unidad Durango, Instituto Politécnico Nacional, Durango, Mexico

²Conservation Science Group, Department of Zoology, University of Cambridge, Cambridge, UK

³Department of Genetics, Evolution and Environment, Centre for Biodiversity and Environment Research, University College London, London, UK

⁴Centre for the Study of Existential Risk, University of Cambridge, Cambridge, UK

⁵School of Biological Sciences, University of Queensland, Brisbane, Qld, Australia

⁶Institute of Zoology, Zoological Society of London, London, UK

Correspondence

Veronica Zamora-Gutierrez, CONACYT – Centro Interdisciplinario de Investigación para el Desarrollo Integral Regional (CIIDIR) Unidad Durango, Instituto Politécnico Nacional, Calle Sigma 119, Fraccionamiento 20 de Noviembre II, Durango 34220, Mexico.

Email: zamora.gtz@gmail.com

Funding information

Whitmore Trust Cambridge; Idea Wild; European Commission's Marie Curie International Incoming Fellowship Program, Grant/Award Number: PIIF-GA-2011-303221; Consejo Nacional de Ciencia y Tecnología, Grant/Award Number: 310731; Rufford Foundation, Grant/Award Number: 12059-1; Isaac Newton Trust; American Society of Mammalogists; Cambridge Commonwealth European and International Trust, Grant/Award Number: 301879989; Hitchcock funds Cambridge; Kenneth Miller Trust; Bat Conservation International; Grantham Foundation for the Protection of the Environment; Engineering and Physical Sciences Research Council, Grant/Award Number: EP/K015664/1

Abstract

Biases in data availability have serious consequences on scientific inferences that can be derived. The potential consequences of these biases could be more detrimental in the less-studied megadiverse regions, often characterized by high biodiversity and serious risks of human threats, as conservation and management actions could be misdirected. Here, focusing on 134 bat species in Mexico, we analyze spatial and taxonomic biases and their drivers in occurrence data; and identify priority areas for further data collection which are currently under-sampled or at future environmental risk. We collated a comprehensive database of 26,192 presence-only bat records in Mexico to characterize taxonomic and spatial biases and relate them to species' characteristics (range size and foraging behavior). Next, we examined variables related to accessibility, species richness and security to explain the spatial patterns in occurrence records. Finally, we compared the spatial distributions of existing data and future threats to these species to highlight those regions that are likely to experience an increased level of threats but are currently under-surveyed. We found taxonomic biases, where species with wider geographical ranges and narrow-space foragers (species easily captured with traditional methods), had more occurrence data. There was a significant oversampling toward tropical regions, and the presence and number of records was positively associated with areas of high topographic heterogeneity, road density, urban, and protected areas, and negatively associated with areas which were predicted to have future increases in temperature and precipitation. Sampling efforts for Mexican bats appear to have focused disproportionately on easily captured species, tropical regions, areas of high species richness and security; leading to under-sampling in areas of high future threats. These biases could substantially influence the assessment of current status of, and future anthropogenic impacts on, this diverse species group in a tropical megadiverse country.

KEYWORDS

anthropogenic threats, chiroptera, data gaps, Mexico, risk hotspots

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

A rapid accumulation of species occurrence data over the past several decades have enabled researchers to carry out large-scale and multitaxa studies to develop methods for inferring present and future species distributions (Soberón & Peterson, 2004). Studies using species occurrence data provide a crucial basis for identifying areas of conservation concern, such as those with high species richness (Platts et al., 2010) and endemic or rare species (Guisan et al., 2006), evaluating the risk of invasive species and diseases (Peterson, 2003), and assessing predictors of species occurrence (Peterman, Crawford, & Kuhns, 2013). Such knowledge can then be used for identifying threats as well as prioritizing and directing future conservation efforts in this anthropozoic era (Groves et al., 2002; Peterman et al., 2013).

Ideally, species occurrence data should be uniform over space and different taxa to reflect the real variation in the distribution, richness, and abundance patterns of species. However, species occurrence data are often seriously biased toward certain regions, habitats, and environmental domains with low coverage particularly in the most biodiverse regions (Hortal, Lobo, & Jiménez-Valverde, 2007; Soberón, Davila, & Golubov, 2004; Yesson et al., 2007). Biases in data availability have serious consequences on scientific inferences that can be derived. For example, spatial data biases can result in an uneven coverage of environmental conditions, such as biomes or climatic zones, where a species could occur (Kadmon, Farber, & Danin, 2004; Loiselle et al., 2007). Such biases can lead to deriving erroneous associations of species with environmental variables (Phillips et al., 2009; Yang, Ma, & Kreft, 2013), inferring incorrect

species' absences (Bystrikova, Peregrym, Erkens, Bezsmertna, & Schneider, 2012), misidentifying areas that have been sampled intensively as species-rich (Petřík, Pergl, & Wild, 2010) and building spurious ecological hypothesis and concepts (Hortal et al., 2015). The potential consequences of using spatially biased data could be more detrimental in less-studied regions, often characterized by high biodiversity and serious risks of human threats, as conservation and management actions could be misdirected (Bini, Diniz-Filho, Rangel, Bastos, & Pinto, 2006). Similarly, if threatened species are less surveyed, the uneven distribution of data over species can also lead to the underestimation of species extinction risk in the taxon (Bland, Collen, Orme, & Bielby, 2015).

Uneven sampling in environmental space is also one of the major sources of errors in studies relating species occurrence data to attributes of the environment in order to model species' niches like species distribution models (SDMs) (Kadmon, Farber, & Danin, 2003; Phillips et al., 2009). If sampling across the environmental space is uneven or if some environmental units are not sampled, the resultant model predictions will be truncated in those under-sampled environmental conditions, which may actually be suitable (Hortal & Lobo, 2008). Model transferability is also highly sensitive to the environmental range covered by the data used to develop and parameterize models (Randin et al., 2006). Incomplete environmental response curves predicted by SDMs can be particularly problematic for environmental change studies as the full environmental tolerances of the species need to be modeled in order to accurately predict geographic changes in habitat suitability.

Mexico is a megadiverse country harboring approximately 10% of all extant species (Mittermeier, Goetsch-Mittermeier, & Robles-Gil,



FIGURE 1 Example of the diversity of Mexican bats. (a) *Antrozous pallidus* and (b) *Lasiurus cinereus* belonging to the Vespertilionidae family, (c) *Pteronotus parnellii* from the Mormoopidae family, and (d) *Natalus mexicanus* member of the Natalidae family. Pictures by Veronica Zamora-Gutierrez

1997) and its natural capital faces serious threats caused by deforestation, species extinctions, and other environmental changes (Malcolm, Liu, Neilson, Hansen, & Hannah, 2006; Mas et al., 2004; Visconti et al., 2011). Bats include a large fraction of the richness and abundances of the Mexican mammalian faunas (~26%, Figure 1) (Ceballos & Oliva, 2005; Medellín, Arita, & Sánchez, 2008) and provide essential functions and services in ecosystems like pollination, seed dispersal, and the suppression of insects' populations (Jones, Jacobs, Kunz, Willig, & Racey, 2009). Bats are considered an ideal indicator taxon for monitoring the status, changes, and responses of biodiversity and ecosystems since their communities respond to changes in temperature and precipitation regimes, farmland expansion, and deforestation (Jones et al., 2009). However, the applicability of indicator taxa is reliant on the existence and availability of data. Although Mexico is considered to be a pioneer in biodiversity informatics (Soberón & Peterson, 2004), there are still considerable knowledge gaps and an uneven distribution of the information available (Sarukhan et al., 2015).

Here, focusing on five biases in the collection and compilation of information (taxonomic, geographical location, species richness, accessibility, and security), we aimed to explain spatial variations in the amount of available bat occurrence data in Mexico and identify priority areas where future survey efforts should be directed. We hypothesized that geographical location, species richness, accessibility, and security could have an impact in explaining spatial biases in available information and that certain groups of species will have more data available. Species occurrence data tend to be more numerous in locations that are easily accessible and less remote (e.g., near cities or with higher road and population densities), with higher actual and/or perceived biodiversity and endemism (e.g., inside protected areas, with complex landscape heterogeneity, and with reported higher species richness) and with high security levels (Amano & Sutherland, 2013; Dennis & Thomas, 2000; Kadmon et al., 2004). Similarly, there tend to be more records for species that are abundant, easily detectable (or captured) and identifiable or charismatic (Hijmans et al., 2000; Meyer et al., 2011; Newbold, 2010).

Our work provides an essential background to inform studies of macroecological patterns and reduce the risk of reaching misleading conclusions in the identification of priority conservation and risk areas. Addressing data gaps and biases is critical to assess the status of biodiversity in megadiverse countries and direct conservation efforts.

2 | MATERIAL AND METHODS

2.1 | Occurrence data

We generated a presence-only database for all bat species that occur in Mexico using information compiled from the National Commission for Knowledge and Use of Biodiversity (CONABIO), Global Biodiversity Information Facility (GBIF), and Mammal Networked Information System (MaNIS). We searched for published literature in English and Spanish using Web of Science, Google Scholar, and Scielo to obtain additional information on Mexican bat occurrence data using the search words "chiroptera," "M*xico," "bat," "bats," "record*," "occurrence," "registro*," "quiropter*," "murci*lago*." Additionally, we requested unpublished material from Mexican researchers. We performed a data cleaning process as described in Zamora-Gutierrez, Pearson, Green, and Jones (2018) leaving 26,192 occurrence records in Mexico (Figure 2) from 134 species (98% of the known species in Mexico) and eight families (100% of the families in Mexico) (Medellín et al., 2008). The total number of grid cells in Mexico was 25,833, of which 9% (2,289 cells) had at least one bat occurrence record (Figure 2). This scarcity of data translates into an area of 1.7 million km² with no bat records.

2.2 | Taxonomic bias

We assessed taxonomic sampling bias based on sampling method and species range size. Hypothetically, the number of occurrence

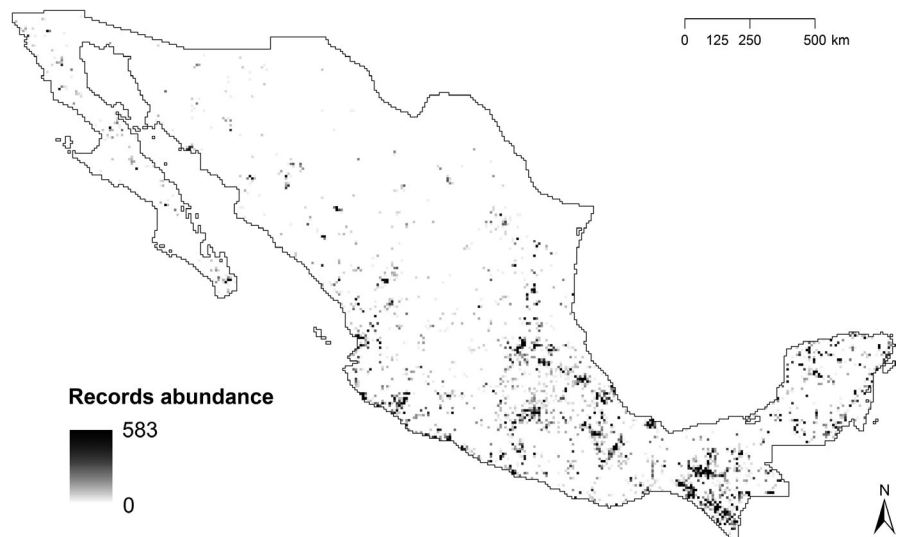


FIGURE 2 Availability of bat occurrence records. The map shows the distribution and abundance of 26,192 bat occurrence records for 134 species (98% of the known species in Mexico) and 8 families (100% of the families in Mexico) across the 2,289 5 arc minute sampled grid cells

data should increase with the size of the area in which a species occurs (Hijmans et al., 2000). To determine range size, we counted the total number of grid cells in which the species is expected to occur using species range maps (IUCN, 2015). Sampling method has been pointed before as a major factor affecting species detectability (Meyer et al., 2011). For example, phyllostomids are highly detectable by frequently used capture methods (e.g., mist nets and harp traps), while insectivorous species are mostly recorded using acoustic methods (MacSwiney, Clarke, & Racey, 2008). We used foraging space as a surrogate for sampling method because such information was not available in the datasets and assigned each species as open-space, edge-space, or narrow-space forager (Denzinger & Schnitzler, 2013).

We used a linear regression to determine the effects of range size (log-transformed and centered) and foraging space (edge-space forager as a reference category) on the number of grid cells sampled (log-transformed) in each species, and included interaction between the two explanatory variables in the model. We then applied a Tukey HSD post hoc test to make a multiple comparison between the foraging groups, using the `glht` function in the `multcomp` package (Hothorn, Bretz, & Westfall, 2008) in R (R Development Core Team, 2013). In the same model, we estimated the slope of log-transformed range size and its 95% confidence interval for each foraging space group using the `cld` function in the `lsmeans` package (Lenth & Love, 2016) in R. This slope can be interpreted as the exponent b of a power-law relationship: $y = a x^b$, where y represents the number of grid cells sampled and x range size (i.e., the number of grid cells in which the species is expected to occur). When the estimated exponent b exceeds 1, it indicates that the proportions of grid cells sampled ($y/x = a x^{b-1}$) are high in wide-ranging species, whereas an exponent smaller than 1 is a sign that the proportions are higher in narrow-ranging species.

2.3 | Spatial bias and its predictors

We first estimated the distribution of all bat occurrence records to assess whether the environmental space and the bat species in Mexico were sampled in an unbiased way. Ideally, we should also use information on abundance and species absence (i.e., a particular location was surveyed but no bat species were detected). However, neither of these types of data is usually available when collating occurrence from these repositories. We thus used the occurrence records of all bat species together as a proxy for the total bat sampling effort across Mexico. We created a grid with the extent of Mexico at a 5 arc minutes resolution ($\sim 10 \text{ km}^2$) and summed the number of bat records occurring in each grid cell and then classified them as either “sampled” (with at least one record of any species) or “un-sampled” (without any record).

To illustrate the geographic variation in sampling activity, we created a map showing a network of Thiessen polygons based on the centroids of all sampled grid cells. This is a scale-independent way of illustrating variation in sampling activity where a big polygon

indicates regionally low sampling activity, whereas a small polygon indicates the opposite (Schulman, Toivonen, & Ruokolainen, 2007; Vale & Jenkins, 2012). This analysis was done using the “Create Thiessen Polygons” tool in ArcGIS v 10.3.

Then, based on the grid cells representing the sampling effort, we assessed how well the environmental conditions in Mexico were sampled by quantifying the spatial sampling bias with respect to the main ecoregions in the country following the formula developed by Kadmon et al., (2004) (Equation 1). The use of ecoregions allowed us to define geographic units that share similar environmental and ecological conditions and are recognized units for global conservation prioritization (INEGI, CONABIO, & INE, 2008).

$$\text{Bias}_d = \frac{n_d - p_d N}{\sqrt{p_d(1 - p_d)N}} \quad (1)$$

where n_d is the number of grid cells with bat records in the ecoregion d , p_d is the proportion of the grid cells in Mexico which fall in ecoregion d , and N is the total number of sampled grid cells ($N = 2,289$). To estimate p , a set of N grid cells (with N equal to the number of actual sampled grid cells) was distributed at random in Mexico, and then p was taken to be the fraction of the random grid cells within a particular ecoregion. To estimate the number of random grid cells, we generated 100 sets of random grid cells of the same number as the sampled grid cells and calculated the mean across the 100 runs per ecoregion. The number of sampled grid cells by ecoregion was then compared with number of random grid cells using a chi-squared test to assess whether sampled grid cells fell into significantly ($\alpha = .05$) different proportions of ecoregions than expected by chance.

We analyzed relationships between the availability of records and seven explanatory variables related to species richness, accessibility, and security level at a resolution of $\sim 10 \text{ km}^2$ grid cells. Three explanatory variables represent species richness: (a) bat species richness, that is, the number of all species that are expected to occur in each grid cell, estimated by overlaying the IUCN range maps of all the bat species distributed in Mexico (IUCN, 2015) and counting how many coincide in each cell, (b) elevational heterogeneity was estimated as the difference in elevation within each grid cell based on a digital elevation model at 60 m resolution (INEGI, 2017), and (c) the coverage of federal protected areas estimated as the percentage of the protected areas polygons within each grid cell (CONANP, 2016). Other three variables represent the accessibility of each cell: (a) road density estimated as the sum within each grid cell of roads length in km (INEGI, 2015), (b) percentage of urban areas estimated as the percentage of the city polygons within each grid cell (INEGI, 2014), and (c) human population density by municipality estimated by assigning the value or average of values given by the map polygons that coincide in each cell (CONABIO, 2010). The remaining variable measures the security level of each grid cell, using reports of the Centro Nacional de Información, which aggregates information on average rate of homicides between 1996 and 2000 per state (SEGOB, 2016). All variables had a correlation less than 0.6 or -0.6 .

To address spatial autocorrelation in residuals, we used conditional autoregressive (CAR) models in this analysis. Considering the zero-inflated nature of the data (i.e., 23,544 out of 25,833 grid cells have no records), we used two response variables in separate regression analyses: (a) the presence of records in each grid cell ($n = 25,833$) and (b) the number of records in each of the grid cell with records ($n = 2,289$). In both models, the explanatory variables were the seven variables described above. All explanatory variables were standardized (mean = 0 and $SD = 1$) before model fitting. We fitted the models to the data with the Markov chain Monte Carlo (MCMC) method in OpenBUGS 3.2.3 (Spiegelhalter, Thomas, Best, & Lunn, 2014) and the R2OpenBUGS package (Gelman et al., 2017) in R (R Development Core Team, 2013). Prior distributions of parameters were set as noninformatively as possible, so as to produce estimates similar to those generated by a maximum likelihood method. We used normal distributions with mean of 0 and variance of 1,000 for coefficients of explanatory variables, an improper uniform distribution (i.e., a uniform distribution on an infinite interval) for the intercept as recommended by Thomas, Best, Lunn, Arnold, and Spiegelhalter (2004), and a Gamma distribution with mean of 1 and variance of 1,000 for the inverses of variance in an intrinsic Gaussian

CAR distribution. Each MCMC algorithm was run with three chains with different initial values for 100,000 iterations with the first 50,000 discarded as burn-in and the remainder thinned to one in every 10 iterations to save storage space. Model convergence was checked with R-hat values (Gelman, Carlin, Stern, & Rubin, 2003).

2.4 | Priority survey areas

We conducted univariate analyses for the relationship between the presence of records in each grid cell ($n = 25,833$) and each of the four threat variables (i.e., predicted changes by 2050s in mean temperature and precipitation, forest and farmland cover percentage per grid cell) using CAR models to address spatial autocorrelation in model residuals, which is highly expected in this type of data. To generate the risk variables, we obtained mean values of annual precipitation and annual temperature at a 5 arc minutes resolution (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005). For future conditions, we selected the CCSM4 General Circulation Model and an extreme greenhouse gas concentration trajectory (Representative Concentration Pathways–RCP) for 2050s with CO_2 predicted to reach around

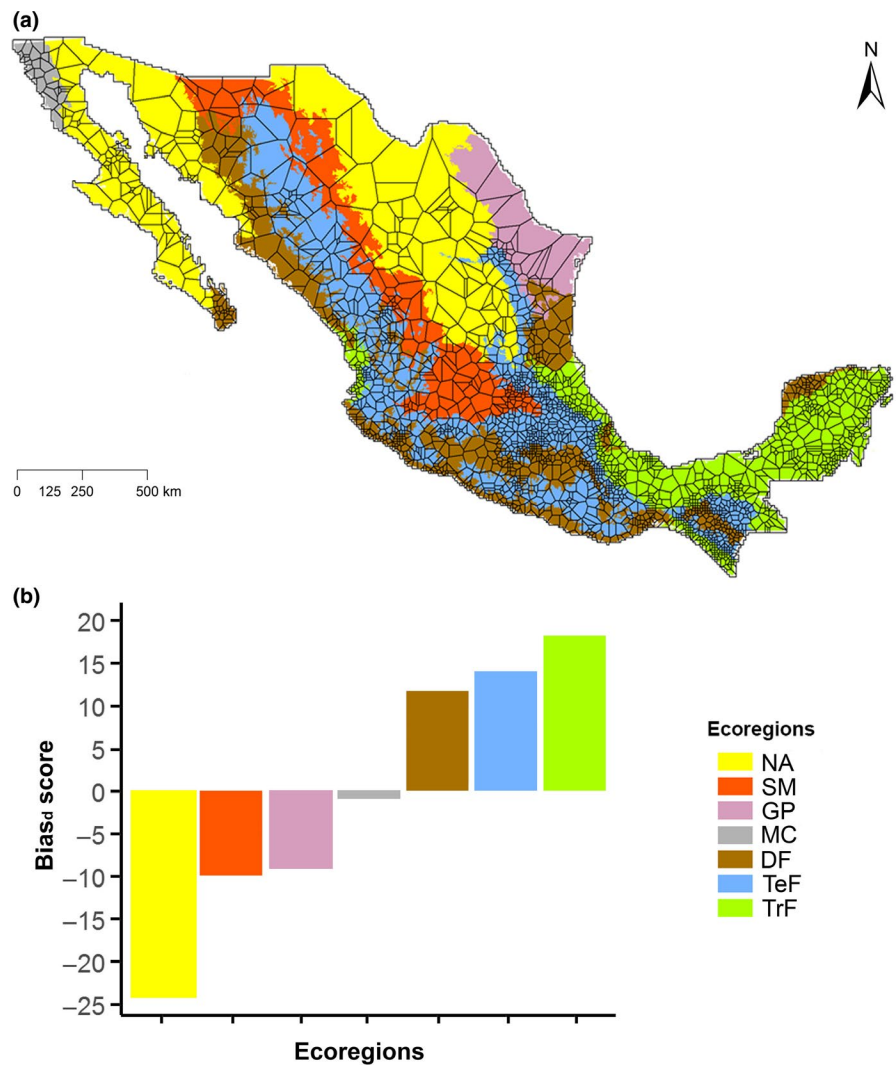


FIGURE 3 Spatial and environmental representation of sampling bias. (a) Sampling activity as the area of Thiessen polygons centered on the centroid of each sampled grid cell. Each polygon shows the area closest to the sampled centroid that is situated in the center of the polygon. The smaller the polygons, the higher the density of sampled grid cells. Colors represent the ecoregions at level I defined by INEGI, CONABIO, & INE, (2008). (b) Environmental bias (Bias_d) in sampled grid cells for Mexico across ecoregions. Bias values go from positive values which indicate the over-sampling, to negative values which indicate environmental under-sampling. DF, Deciduous Forest; GP, Great Plains; MC, Mediterranean California; NA, North American Deserts; SM, Semiarid Meridional Elevations; TeF, Temperate Forest; TrF, Tropical Forest

TABLE 1 Results of the linear model comparing the effects of range size (log-transformed and centered) and foraging space (edge-space forager as a reference category) and their interaction on the number of grid cells sampled (log-transformed) in each species (Model 1)

	Model 1			Model 2				Model 3	
	df	F	p	Estimate	SE	t value	p _{adj}	b	95% CI
Range size	1,128	177.80	<.001						
Foraging space	2,128	12.78	<.001						
Interaction	2,128	3.28	<.04						
N-E				0.63	0.20	3.22	0.004		
O-N				-1.03	0.23	4.49	0.001		
O-E				-0.40	0.25	-1.63	0.23		
O								0.53	0.26–0.79
E								0.71	0.55–0.86
N								0.94	0.74–1.15

Note: Model 2 shows the results of the Tukey HSD post hoc test for a multiple comparison between the foraging groups (O = open-space foragers, E = edge-space foragers, N = narrow-space foragers). Model 3 summarizes the results of Model 2 with an additional estimate of the slope of log-transformed range size and its 95% confidence interval for each foraging space group. This slope can be interpreted as the exponent b of a power-law relationship: $y = a x^b$, where y represents the number of grid cells sampled and x range size (i.e., the number of grid cells in which the species is expected to occur).

2,000 ppmv (RCP-8.5) (IPCC, 2013). Current and future land use maps were obtained from van Eupen et al., (2014). For future land conditions, we selected two extreme socioeconomic contexts: (a) a sustainable scenario where the use of resources and fossil fuels and deforestation is reduced; and (b) a “business-as-usual” scenario where no efforts to reduced land degradation and the use of fossil fuels are taken. We used four land use classes (forest, shrubland, grassland, and cropland) at 5 arc minute resolution and estimated their proportions per each grid cell.

3 | RESULTS

3.1 | Sampling biases

Thiessen polygons network showed that the knowledge of the bat fauna in the North American deserts and the great plains is based on collections from a few widely spread sampled grid cells (Figure 3a). The area of the largest polygon is more than 48,000 km², indicating that knowledge of bats in this area is based upon a single sampled grid cell. There were 31 of these large polygons ($\geq 10,000$ km²) covering more than 59 million km², 16 of which were located in the North American desert ecosystem, six in the great plains, four in the north of the western mountain range and five in the north of the semiarid meridional elevations. Supporting this result, the proportion of sampled grid cells in different ecoregions was significantly different from that expected by chance ($\chi^2 = 645.64$, $df = 6$, $p < .001$). The calculation of Bias_d revealed that three ecoregions, North American deserts, semiarid meridional elevations and great plains, have particularly been under-sampled while three ecoregions, tropical forest, temperate forest, and deciduous forest, have been over-sampled (Figure 3b).

The number of reported species per sampled grid cell was very low with 68% of the sampled grid cells having ≤ 5 species reported.

The percentage of all grid cells within a species' range that were sampled ranged between 0.1% and 27%, with 50 (37%) species having ≤ 20 grid cells sampled. There was a significant effect of range size, foraging space, and their interaction on the number of sampled grid cells (Table 1, Model 1). Post hoc comparison showed that number of sampled grid cells for narrow-space foragers was, when controlled for their range size, significantly higher than those for edge- and open-space foragers, while there was no significant difference between edge- and open-space foragers (Table 1, Model 2). The estimated exponent b in the relationship $y = a x^b$ (y = the number of grid cells sampled, x = range size) was significantly smaller than 1 for open-space foragers and for edge-space foragers, while it was not significantly different from 1 for narrow-space foragers (Table 1, Model 3). This indicates that the proportions of grid cells sampled were higher in narrow-ranging species for open- and edge-space foragers.

3.2 | Predictors of spatial bias

Results of CAR models indicate that both the attractiveness and accessibility of grid cells are important predictors of data availability. The presence of records in each grid cell was associated particularly with high topographic heterogeneity, road density, percentages of city area, and protected areas (Figure 4a). Similarly, among grid cells with records, there were more records in grid cells with higher species richness, road density, and percentages of city areas and protected areas (Figure 4b).

3.3 | Priority survey areas

The presence of records was negatively associated with changes in temperature and precipitation (Figure 5), indicating that records are

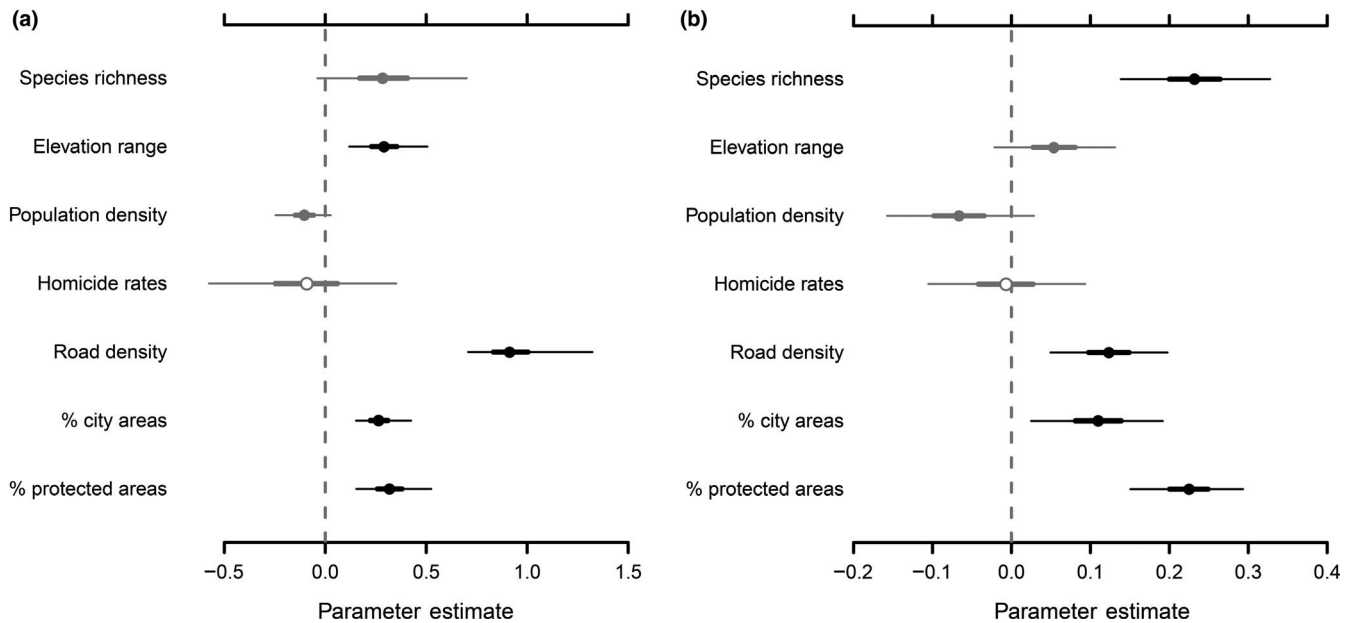


FIGURE 4 Effects of hypothesized explanatory variables on (a) the presence of records in each grid cell ($n = 25,833$); and (b) the number of records in each of the grid cell with records ($n = 2,289$). Estimated coefficients (posterior medians) with 95% and 50% (thick lines) credible intervals are shown (those not overlapping with zero shown with filled circles in black)

particularly scarce in areas with predicted increases in temperature and precipitation. On the other hand, the presence of records was associated with neither forest change nor farmland change (Figure 5). Nevertheless this means that areas with predicted deforestation

and farmland area expansion are not necessarily surveyed well compared to other areas. Priority areas for further surveys which are currently under-sampled and at higher risk of environmental change include the North American Deserts, Semiarid Meridional Elevations and the Great Plains.

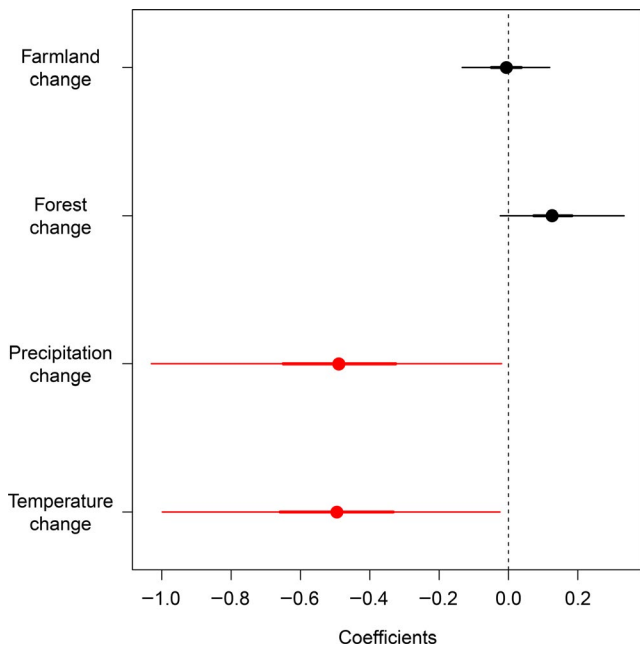


FIGURE 5 Estimated coefficients in univariate analyses testing the relationship between the presence of records and each of the four threat variables: predicted changes in temperature, precipitation, forest cover and farmland area. Posterior medians with 95% and 50% (thick lines) credible intervals are shown. Coefficients with 95% credible intervals not overlapping with zero are shown in red

4 | DISCUSSION

Although the dataset collated here represents the most comprehensive set of occurrence records for Mexican bats, it shows that Mexico is poorly sampled with 90% of the territory with no bat occurrence records and with a sampling effort unevenly distributed across the country. The substantial lack of biodiversity data coverage in megadiverse regions has been pointed out in previous studies (Amano & Sutherland, 2013; Bernard, Aguiar, & Machado, 2011; Collen, Ram, Zamin, & McRae, 2008; Llorente-Bousquets, Ocegueda, Contreras, Chiang, & Papavero, 2008) but this is the first time such detailed analysis has been done for Mexican bats. The sampling gaps found in Mexico are a matter of concern as they could substantially influence the way ecological and biodiversity value of different parts of the country are perceived. If patterns and changes in biodiversity cannot be measured accurately, threats to species and ecosystems cannot be adequately addressed and consequently conservation efforts might be misdirected (Collen et al., 2008).

4.1 | Data reliability

Spatial sampling bias over different ecoregions was evident in the existing bat occurrence records in Mexico. Sampling appears to

have focused disproportionately on the tropical regions, while the bat fauna of the drylands are represented by only a few scattered sampled grid cells. Hence, the completeness of the knowledge of the Nearctic and the Neotropical bat faunas is unlikely to be equal. This pattern of over and under-sampling of some ecoregions has been previously identified in other temperate and arid regions (Bernard et al., 2011; Martin, Blossey, & Ellis, 2012; Meyer, Kreft, Guralnick, & Jetz, 2015). Bats are crucial components of these under-sampled environments as they provide key ecosystem services and have high mammalian richness and abundance (Carpenter, 1969; Jones et al., 2009). Thus, the uneven coverage in the knowledge of the bat faunas may have important consequences in safeguarding the natural capital of these regions and the provision of services to the human well-being (e.g., biological insect pest suppression).

Mexico has the highest percentage of bat occurrence records from Latin America (37%) in the GBIF portal (Noguera-Urbano & Escalante, 2014). Yet, more than 50% of the sampled grid cells have five or less species recorded and none of the species had more than 30% of its range sampled. Despite the large efforts by Mexican institutions to increase the availability of electronic verified records (Soberón & Peterson, 2004), data quantities for a large number of species are still insufficient. Although the efforts from different sources and researchers have allowed us to achieve some understanding of the distribution of richness and species for Mexican bats, this picture might not be a fair representation of the true distributions (Sastre & Lobo, 2009). When species richness is predicted using biased data, the predicted values tend to be higher in the most surveyed localities compared to those under-sampled. Hence, biased sampling effort can also bias estimates of species richness (Petřík et al., 2010; Sánchez-Fernández, Lobo, Abellán, Ribera, & Millán, 2008).

The analysis presented here shows that the representation of bat species in the occurrence records was influenced not only by species' range size but also by their foraging behavior. The effect of foraging style on the amount of sampled grid cells could be explained by two main reasons: (a) the over-sampled tropical ecosystems might have been dominated by narrow-space foragers (Fenton et al., 1992); or (b) narrow-space foragers are the best detected group by capture methods (mist nets) which are preferably used in the Neotropics (Barnett et al., 2006). Mist nets undoubtedly favor the detection of some narrow-space foragers, such as phyllostomid species and under-sample open- and edge-space foragers (e.g., molossids and vespertilionids) (MacSwiney et al., 2008). However, recent technological innovations have led to the increased use of new survey methods (e.g., ultrasonic detectors) (Jones et al., 2013) that could change the overall representativeness of species sampled over time. However, changes in species detectability caused by changes in survey methods may lead to erroneous assumptions about species distributional changes. It is important to further explore effects of survey methods on species data representativeness in order to disentangle changes attributable to sampling rather than environmental or ecological drivers (see da Rocha, Ferrari, Feijó, & Gouveia, 2015), and to apply correcting methods when dealing with biased data. For example, the

information contained in occurrence records can be improved (without losing data) with some recently developed statistical techniques like random walk priors in Bayesian occupancy models (Outhwaite et al., 2018), and specific target-group background selection methods when building SDMs (Phillips et al., 2009).

4.2 | Drivers of bias

We found that variables that favor high species richness within an area (i.e., protected areas and topographic heterogeneity) had a positive effect on the availability of bat occurrence data. This result is consistent with previous studies where sampling effort is biased toward regions considered as conservation priorities (Parnell et al., 2003; Reddy & Dávalos, 2003; Yang, Ma, & Kreft, 2014). Plausible explanations include that protected areas are more secured, which is a considerable barrier for researchers in countries under conflict (Amano & Sutherland, 2013). Also, protected areas are more likely to host pristine habitats which might increase the chances to find endemic, rare, and a higher number of species. Biases toward areas with perceived high species richness can be particularly problematic for conservation purposes because it can give the wrong impression of certain areas to be species rich, which can lead to misleading prioritization for conservation planning (Boakes, Fuller, McGowan, & Mace, 2016).

The concentration of bat occurrence data in highly heterogeneous habitats like mountains can be explained by three main factors. First, the most important institutions for biodiversity and taxonomic research are located in Mexico city and toward the center and south of the country (Llorente-Bousquets, Michán, et al., 2008; Llorente-Bousquets, Ocegueda, et al., 2008). These institutes are located close to areas of high mammal's biodiversity and endemism like the states of Oaxaca, Chiapas, and Veracruz (Ceballos & Oliva, 2005). Second, these biodiverse states host important and complex mountainous areas in Mexico like the Sierra Madre de Chiapas and Sierra Madre del Sur, and unique vegetation types like the cloud forest. Third, the east and west mountain ranges in the country are considered between having medium to high mammal's diversity (Ceballos & Oliva, 2005). We can thus conclude that the highly biodiverse, pristine, and heterogeneous mountainous areas of Mexico have attracted the attention of trained taxonomist over the years (see references within Ceballos & Oliva, 2005).

Location's accessibility is also an important predictor of data availability for both presence and abundance of records, meaning that more accessible places tend to be more frequently sampled and with more data. These biases in occurrence data could lead to erroneous ecological patterns. For example, previous studies have found that off-road avian community composition significantly differs from samples taken around roads (Wellcome, Kardynal, Franken, & Gillies, 2014). Similar patterns might hold for bats as different species and trophic groups have different tolerance to disturbance and human-made environments (Estrada-Villegas, Meyer, & Kalko, 2010; Gonçalves, Fischer, & Dirzo, 2017; Jung & Kalko, 2011). Sampling

bias toward densely populated areas can cause erroneous conclusions about associations between species and people, and it may underestimate the impact of human activities on biodiversity (Barbosa, Pautasso, & Figueiredo, 2013; Ficetola et al., 2014). Mexico leads the list of countries with the highest deforestation rates (Mas et al., 2004) and with the highest human densities (World Bank, 2017). The biases found here toward anthropogenic elements give caution in making ecological and biogeographic inferences on bat distributions as it might have serious consequences for conservation and management plans. It is important to study data biases involving human-made environments to disentangle sampling artifacts from environmental patterns and associations. Security aspects may also influence field sampling but we did not find evidence that insecure conditions limit the number of records available in Mexico. This result might be scale dependent as we used a security metric given at state level. A more detailed analysis on specific dangerous areas might give a different result as it is known that scientist tend to avoid sampling conflict areas (Amano & Sutherland, 2013; Boakes et al., 2016).

4.3 | Prioritization

This study identified the areas that need to be better explored and characterized in terms of their bat fauna. Our study emphasizes the need to improve bat data, and very likely from other taxa, in areas with specific environmental (drylands), geographic (northern Mexico and outside protected areas), and socio-economic characteristics (e.g., remote away from cities and roads), and for specific bat groups (e.g., open-space foragers) to improve the quality and representativeness of distributional data and consequently species knowledge to take well-informed conservation decisions. The sampling biases identified here lead to an evident Wallacean shortfall, that is, poorly understood and incomplete knowledge on species distributions. It is particularly worrying that the lack of records is found in areas projected to have extreme weather events by 2050s and that are likely to suffer high rates of land use change. This lack of information is likely to limit (a) our understanding on how biodiversity is and will be affected by environmental changes and (b) subsequently our capacity to develop sound conservation strategies.

Ideally, species should be sampled homogeneously across space and time and a good example of long term and long scale biodiversity surveys can be seen at the Butterfly Monitoring Scheme which started in the UK in the mid-1970s, and since then, it has been extended to more than ten European countries. This monitoring program is based on standardized sampling protocols designed to collect data for both common and rare species, covering a wide range of habitats, and it is constantly updated and adjusted to detect and reduce data gaps by prioritizing survey efforts (Van Swaay, Nowicki, Settele, & Strien, 2008). This type of surveys must be implemented in countries highlighted as data deficient and of high conservation priorities like Mexico (e.g., Brazil, Bernard et al., 2011).

The next step is to fill in the information gaps identified here. Mexico is leader in data mobilization through the CONABIO (Soberón & Peterson, 2004); thus, this data biases might be a consequence of the lack of sampling effort, rather than data sharing. A crucial action is to prioritize data collection to minimize efforts and money. Citizen science programs have proven to be effective to increase the representativeness of some taxa on digital occurrence databases (Amano, Laming, & Sutherland, 2016), as they can be done over large regions at a low cost. However, citizen science data collection may continue promoting certain biases like accessibility and security. Overcoming these biases will be particularly challenging in megadiverse regions as the most data deficient regions are less accessible and secure, and require a great investment of time and money to reach (Bernard et al., 2011).

Indirect observation through acoustic surveys is particularly useful for less detectable groups like bats and they have been successfully implemented to improve the coverage of bat's distributional data (Jones et al., 2013). In any case, further sampling must be encouraged toward areas and species that require immediate attention. In other to close the information gap across species and regions, new data collection programs should focus on the regions and species groups with data deficiencies identified here. Survey design should be stratified so that sampling must be systematic and carried out by trained biologists or surveyors.

ACKNOWLEDGEMENTS

This study was financially supported by American Society of Mammalogists, Bat Conservation International, Cambridge Commonwealth European and International Trust (No. 301879989), CONACYT (No. 310731), Hitchcock funds Cambridge, Idea Wild, Rufford Small Grants (No. 12059-1) and Whitmore Trust Cambridge to V.Z.G, the European Commission's Marie Curie International Incoming Fellowship Program (PIIF-GA-2011-303221), Isaac Newton Trust, Grantham Foundation for the Protection of the Environment and the Kenneth Miller Trust to T.A. and Engineering and Physical Sciences Research Council (EPSRC) Grant EP/K015664/1 to K.E.J. We thank all the people who shared their data and to two anonymous reviewers for their valuable comments on earlier versions of the manuscript.

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTIONS

V.Z.G. and K.E.J. conceived and designed the study, V.Z.G. collected the data, V.Z.G. and T.A. analyzed the data, V.Z.G. led the writing with contributions from all authors.

ORCID

Veronica Zamora-Gutierrez  <https://orcid.org/0000-0003-0661-5180>

Tatsuya Amano  <https://orcid.org/0000-0001-6576-3410>

Kate E. Jones  <https://orcid.org/0000-0001-5231-3293>

DATA AVAILABILITY STATEMENT

Data from the variables used for the analysis have been deposited at Dryad: <https://doi.org/10.5061/dryad.qrfj6q5b3>

REFERENCES

- Amano, T., Lamming, J. D. L., & Sutherland, W. J. (2016). Spatial gaps in global biodiversity information and the role of citizen science. *BioScience*, 66, 393–400.
- Amano, T., & Sutherland, W. J. (2013). Four barriers to the global understanding of biodiversity conservation: Wealth, language, geographical location and security. *Proceedings of the Royal Society B: Biological Sciences*, 280, 20122649.
- Barbosa, A. M., Pautasso, M., & Figueiredo, D. (2013). Species – People correlations and the need to account for survey effort in biodiversity analyses. *Diversity and Distributions*, 19, 1188–1197. <https://doi.org/10.1111/ddi.12106>
- Barnett, A. A., Sampaio, E. M., Kalko, E. K. V., Shapley, L., Fischer, E., Camargo, G., & Rodríguez-Herrera, B. (2006). Bats of Jaú National Park, central Amazonia, Brazil. *Acta Chiropterologica*, 8, 103–128.
- Bernard, E., Aguiar, L. M. S., & Machado, R. B. (2011). Discovering the Brazilian bat fauna: A task for two centuries? *Mammal Review*, 41, 23–39. <https://doi.org/10.1111/j.1365-2907.2010.00164.x>
- Bini, L. M., Diniz-Filho, J. A. F., Rangel, T. F. L. V. B., Bastos, R. P., & Pinto, M. P. (2006). Challenging Wallacean and Linnean shortfalls: Knowledge gradients and conservation planning in a biodiversity hotspot. *Diversity and Distributions*, 12, 475–482. <https://doi.org/10.1111/j.1366-9516.2006.00286.x>
- Bland, L. M., Collen, B., Orme, C. D. L., & Bielby, J. (2015). Predicting the conservation status of data-deficient species. *Conservation Biology*, 29, 250–259. <https://doi.org/10.1111/cobi.12372>
- Boakes, E. H., Fuller, R. A., McGowan, P. J. K., & Mace, G. M. (2016). Uncertainty in identifying local extinctions: The distribution of missing data and its effects on biodiversity measures. *Biology Letters*, 12, 1–4.
- Bystrikova, N., Peregrym, M., Erkens, R. H. J., Bezsmertna, O., & Schneider, H. (2012). Sampling bias in geographic and environmental space and its effect on the predictive power of species distribution models. *Systematics and Biodiversity*, 10, 305–315. <https://doi.org/10.1080/14772000.2012.705357>
- Carpenter, R. E. (1969). Structure and function of the kidney and the water balance of desert bats. *Physiological Zoology*, 42, 288–302. <https://doi.org/10.1086/physzool.42.3.30155492>
- Ceballos, G., & Oliva, G. (2005). *Los mamíferos silvestres de México*. Mexico City, Mexico: FCE, CONABIO.
- Collen, B., Ram, M., Zamin, T., & McRae, L. (2008). The tropical biodiversity data gap: Addressing disparity in global monitoring. *Tropical Conservation Science*, 1, 75–88. <https://doi.org/10.1177/194008290800100202>
- CONABIO (2010). Retrieved from http://www.conabio.gob.mx/informacion/metadatos/gis/dipomun00gw.xml?_xsl=/db/metadatos/xsl/fgdc_html.xsl&indent=no
- CONANP (2016). Retrieved from http://sig.conanp.gob.mx/website/pagsig/info_shape.htm
- da Rocha, P. A., Ferrari, S. F., Feijó, A., & Gouveia, S. F. (2015). Zoogeography of South American forest-dwelling bats: Disjunct distributions or sampling deficiencies? *PLoS ONE*, 10, e0133276. <https://doi.org/10.1371/journal.pone.0133276>
- Dennis, R. L. H., & Thomas, C. D. (2000). Bias in butterfly distribution maps: The influence of hot spots and recorder's home range. *Journal of Insect Conservation*, 4, 73–77.
- Denzinger, A., & Schnitzler, H.-U. (2013). Bat guilds, a concept to classify the highly diverse foraging and echolocation behaviors of microchiropteran bats. *Frontiers in Physiology*, 4, 1–15. <https://doi.org/10.3389/fphys.2013.00164>
- Estrada-Villegas, S., Meyer, C. F. J., & Kalko, E. K. V. (2010). Effects of tropical forest fragmentation on aerial insectivorous bats in a land-bridge island system. *Biological Conservation*, 143, 597–608. <https://doi.org/10.1016/j.biocon.2009.11.009>
- Fenton, M. B., Acharya, L., Audet, D., Hickey, M. B. C., Merriman, C., Obrist, M. K., ... Adkins, B. (1992). Phyllostomid bats (Chiroptera: Phyllostomidae) as indicators of habitat disruption in the Neotropics. *Biotropica*, 24, 440–446. <https://doi.org/10.2307/2388615>
- Ficetola, G. F., Cagnetta, M., Padoa-schioppa, E., Quas, A., Razzetti, E., Sindaco, R., & Bonardi, A. (2014). Sampling bias inverts ecogeographical relationships in island reptiles. *Global Ecology and Biogeography*, 23, 1303–1313. <https://doi.org/10.1111/geb.12201>
- Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2003). *Bayesian data analysis*. Boca Raton, FL: Chapman & Hall/CRC.
- Gelman, A., Sturtz, S., Ligges, U., Gorjanc, G., Kerman, J., & Thomas, N. (2017). Retrieved from <https://cran.r-project.org/web/packages/R2OpenBUGS/index.html>
- Gonçalves, F., Fischer, E., & Dirzo, R. (2017). Forest conversion to cattle ranching differentially affects taxonomic and functional groups of Neotropical bats. *Biological Conservation*, 210, 343–348. <https://doi.org/10.1016/j.biocon.2017.04.021>
- Groves, C. R., Jensen, D. B., Valutis, L. L., Redford, K. H., Shaffer, M. L., Scott, J. M., ... Anderson, M. G. (2002). Planning for biodiversity conservation: Putting conservation science into practice. *BioScience*, 52, 499. [https://doi.org/10.1641/0006-3568\(2002\)052\[0499:PFBCPC\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2002)052[0499:PFBCPC]2.0.CO;2)
- Guisan, A., Broennimann, O., Engler, R., Vust, M., Yoccoz, N. G., Lehmann, A., & Zimmermann, N. E. (2006). Using niche-based models to improve the sampling of rare species. *Conservation Biology*, 20, 501–511. <https://doi.org/10.1111/j.1523-1739.2006.00354.x>
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965–1978. <https://doi.org/10.1002/joc.1276>
- Hijmans, R. J., Garrett, K. A., Huaman, Z., Zhang, D. P., Schreuder, M., & Bonierbale, M. (2000). Assessing the geographic representativeness of genebank collections: The case of Bolivian wild potatoes. *Conservation Biology*, 14, 1755–1765. <https://doi.org/10.1111/j.1523-1739.2000.98543.x>
- 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, 523–549. <https://doi.org/10.1146/annurev-ecolsys-112414-054400>
- Hortal, J., Jiménez-Valverde, A., Gómez, J. F., Lobo, J. M., & Baselga, A. (2008). Historical bias in biodiversity inventories affects the observed environmental niche of the species. *Oikos*, 117, 847–858. <https://doi.org/10.1111/j.0030-1299.2008.16434.x>
- Hortal, J., Lobo, J. M., & Jiménez-Valverde, A. (2007). Limitations of biodiversity databases: Case study on seed-plant diversity in Tenerife, Canary Islands. *Conservation Biology*, 21, 853–863. <https://doi.org/10.1111/j.1523-1739.2007.00686.x>
- Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in General Parametric Models. *Biometrical Journal*, 50, 346–363. <https://doi.org/10.1002/bimj.200810425>
- Hothorn, T., Bretz, F., Westfall, P., Heiberger, R., Schuetzenmeister, A., & Scheibe, S. (2017). *Simultaneous inference in general parametric models*. 36.
- INEGI (2014). Retrieved from http://www.inegi.org.mx/geo/contenidos/geoestadistica/m_g_0.aspx
- INEGI (2015). Retrieved from <http://www3.inegi.org.mx/sistemas/biblioteca/ficha.aspx?upc=702825209575>
- INEGI (2017). Retrieved from <http://www.inegi.org.mx/geo/contenidos/datosrelieve/continental/Descarga.aspx>

- INEGI, CONABIO, & INE (2008). Retrieved from http://www.conabio.gob.mx/informacion/metadatos/gis/ecort08gw.xml?_xsl=/db/metadatos/xsl/fgdc_html.xsl&_indent=no
- IPCC (2013). *Climate Change 2013: The physical science basis. Contribution of Working Group I to the fifth assessment report of the intergovernmental panel on climate change*. Cambridge, UK and New York, NY: Cambridge University Press.
- IUCN (2015). Retrieved from <http://www.iucnredlist.org>
- Jones, G., Jacobs, D. S., Kunz, T. H., Willig, M. R., & Racey, P. A. (2009). Carpe noctem: The importance of bats as bioindicators. *Endangered Species Research*, 8, 93–115. <https://doi.org/10.3354/esr00182>
- Jones, K. E., Russ, J. A., Bashta, A.-T.-T., Bilhari, Z., Catto, C., Csósz, I., ... Zavarzin, O. (2013). Indicator bats program: A system for the global acoustic monitoring of bats. In B. Collen, N. Pettorelli, J. E. M. Baillie, & S. Durant (Eds.), *Biodiversity monitoring and conservation: Bridging the gaps between global commitment and local action* (p. 31). London, UK: Wiley & Sons.
- Jung, K., & Kalko, E. K. V. (2011). Adaptability and vulnerability of high flying Neotropical aerial insectivorous bats to urbanization. *Diversity and Distributions*, 17, 262–274. <https://doi.org/10.1111/j.1472-4642.2010.00738.x>
- Kadmon, R., Farber, O., & Danin, A. (2003). A systematic analysis of factors affecting the performance of climatic envelope models. *Ecological Applications*, 13, 853–867. [https://doi.org/10.1890/1051-0761\(2003\)013\[0853:ASA0FA\]2.0.CO;2](https://doi.org/10.1890/1051-0761(2003)013[0853:ASA0FA]2.0.CO;2)
- Kadmon, R., Farber, O., & Danin, A. (2004). Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecological Applications*, 14, 401–413. <https://doi.org/10.1890/02-5364>
- Lenth, R., & Love, J. (2016). Least-Squares Means: The R package lsmeans. *Journal of Statistical Software*, 69, 1–33. <https://doi.org/10.18637/jss.v069.i01>
- Llorente-Bousquets, J., Michán, L., González-González, J., Sosa-Ortega, V., Espinosa-Organista, D., Hoffman, A., & Ramírez-Pulido, J. (2008). Desarrollo y situación del conocimiento de las especies. In J. Soberón, G. Halfter, & J. Llorente-Bousquets (Eds.), *Capital natural de México* (Vol. I, pp. 193–214). Mexico City, Mexico: CONABIO.
- Llorente-Bousquets, J., Ocegueda, S., Contreras, S., Chiang, F., & Papavero, N. (2008). Estado del conocimiento de la biota. In J. Soberón, G. Halfter, & J. Llorente-Bousquets (Eds.), *Capital natural de México, vol. I: Conocimiento actual de la biodiversidad* (pp. 283–322). Mexico City, Mexico: CONABIO.
- Loiselle, B. A., Jørgensen, P. M., Consiglio, T., Jiménez, I., Blake, J. G., Lohmann, L. G., & Montiel, O. M. (2007). Predicting species distributions from herbarium collections: Does climate bias in collection sampling influence model outcomes? *Journal of Biogeography*, 35, 105–116. <https://doi.org/10.1111/j.1365-2699.2007.01779.x>
- MacSwiney, M. C., Clarke, F. M., & Racey, P. A. (2008). What you see is not what you get: The role of ultrasonic detectors in increasing inventory completeness in Neotropical bat assemblages. *Journal of Applied Ecology*, 45, 1364–1371. <https://doi.org/10.1111/j.1365-2664.2008.01531.x>
- Malcolm, J. R., Liu, C., Neilson, R. P., Hansen, L., & Hannah, L. (2006). Global warming and extinctions of endemic species from biodiversity hotspots. *Conservation Biology*, 20, 538–548. <https://doi.org/10.1111/j.1523-1739.2006.00364.x>
- Martin, L. J., Blossey, B., & Ellis, E. (2012). Mapping where ecologists work: Biases in the global distribution of terrestrial ecological observations. *Frontiers in Ecology and the Environment*, 10, 195–201. <https://doi.org/10.1890/110154>
- Mas, J., Velázquez, A., Díaz-Gallegos, J. R., Mayorga-Alcántara, R., Alcántara, C., Bocco, G., ... Pérez-Vega, A. (2004). Assessing land use/cover changes: A nationwide multirate spatial database for Mexico. *International Journal of Applied Earth Observation and Geoinformation*, 5, 249–261. <https://doi.org/10.1016/j.jag.2004.06.002>
- Medellín, R. A., Arita, H. T., & Sánchez, O. D. F. (2008). *Identificación de los murciélagos de México, clave de campo*. Mexico, D.F., Mexico: Instituto de Ecología, UNAM – CONABIO
- Meyer, C. F. J., Aguiar, L. M. S., Aguirre, L. F., Baumgarten, J., Clarke, F. M., Cosson, J.-F., ... Kalko, E. K. V. (2011). Accounting for detectability improves estimates of species richness in tropical bat surveys. *Journal of Applied Ecology*, 48, 777–787. <https://doi.org/10.1111/j.1365-2664.2011.01976.x>
- Meyer, C., Kreft, H., Guralnick, R., & Jetz, W. (2015). Global priorities for an effective information basis of biodiversity distributions. *Nature Communications*, 6, 8221. <https://doi.org/10.1038/ncomms9221>
- Mittermeier, R. A., Goetsch-Mittermeier, C., & Robles-Gil, P. (1997). *Megadiversity: Earth's biologically wealthiest nations*. Mexico, D.F., Mexico: Cemex.
- Newbold, T. (2010). Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. *Progress in Physical Geography*, 34, 3–22. <https://doi.org/10.1177/0309133309355630>
- Noguera-Urbano, E. A., & Escalante, T. (2014). Datos geográficos de los murciélagos (Chiroptera) en el Neotrópico. *Revista De Biología Tropical*, 62, 201–215. <https://doi.org/10.15517/rbt.v62i1.8379>
- Outhwaite, C. L., Chandler, R. E., Powney, G. D., Collen, B., Gregory, R. D., & Isaac, N. J. B. (2018). Prior specification in Bayesian occupancy modelling improves analysis of species occurrence data. *Ecological Indicators*, 93, 333–343. <https://doi.org/10.1016/j.ecoli.2018.05.010>
- Parnell, J. A. N., Simpson, D. A., Moat, J., Kirkup, D. W., Chantaranonthai, P., Boyce, P. C., ... Wilkin, P. (2003). Plant collecting spread and densities: Their potential impact on biogeographical studies in Thailand. *Journal of Biogeography*, 30, 193–209. <https://doi.org/10.1046/j.1365-2699.2003.00828.x>
- Peterman, W. E., Crawford, J. A., & Kuhns, A. R. (2013). Using species distribution and occupancy modeling to guide survey efforts and assess species status. *Journal for Nature Conservation*, 21, 114–121. <https://doi.org/10.1016/j.jnc.2012.11.005>
- Peterson, A. T. (2003). Predicting the geography of species' invasions via ecological niche modeling. *The Quarterly Review of Biology*, 78, 419–433. <https://doi.org/10.1086/378926>
- Petrík, P., Pergl, J., & Wild, J. (2010). Recording effort biases the species richness cited in plant distribution atlases. *Perspectives in Plant Ecology, Evolution and Systematics*, 12, 57–65. <https://doi.org/10.1016/j.ppees.2009.06.004>
- Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecological Applications*, 19, 181–197. <https://doi.org/10.1890/07-2153.1>
- Platts, P. J., Ahrends, A., Gereau, R. E., McClean, C. J., Lovett, J. C., Marshall, A. R., ... Marchant, R. (2010). Can distribution models help refine inventory-based estimates of conservation priority? A case study in the Eastern Arc forests of Tanzania and Kenya. *Diversity and Distributions*, 16, 628–642. <https://doi.org/10.1111/j.1472-4642.2010.00668.x>
- R Development Core Team (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Development Core Team.
- Randin, C. F., Dirnböck, T., Dullinger, S., Zimmermann, N. E., Zappa, M., & Guisan, A. (2006). Are niche-based species distribution models transferable in space? *Journal of Biogeography*, 33, 1689–1703. <https://doi.org/10.1111/j.1365-2699.2006.01466.x>
- Reddy, S., & Dávalos, L. M. (2003). Geographic sampling bias and its implications for conservation priorities in Africa. *Journal of Biogeography*, 30, 1719–1727.
- Sánchez-Fernández, D., Lobo, J. M., Abellán, P., Ribera, I., & Millán, A. (2008). Bias in freshwater biodiversity sampling: The case of Iberian water beetles. *Diversity and Distributions*, 14, 754–762. <https://doi.org/10.1111/j.1472-4642.2008.00474.x>

- Sarukhan, J., Urquiza-Haas, T., Koleff, P., Carabias, J., Dirzo, R., Ezcurra, E., ... Soberon, J. (2015). Strategic actions to value, conserve, and restore the natural capital of megadiversity countries: The case of Mexico. *BioScience*, 65, 164–173. <https://doi.org/10.1093/biosci/biu195>
- Sastre, P., & Lobo, J. M. (2009). Taxonomist survey biases and the unveiling of biodiversity patterns. *Biological Conservation*, 142, 462–467. <https://doi.org/10.1016/j.biocon.2008.11.002>
- Schulman, L., Toivonen, T., & Ruokolainen, K. (2007). Analysing botanical collecting effort in Amazonia and correcting for it in species range estimation. *Journal of Biogeography*, 34, 1388–1399. <https://doi.org/10.1111/j.1365-2699.2007.01716.x>
- SEGOB (2016). *Tasas por cada 100 mil habitantes 1997–2016*.
- Soberón, J., Davila, P., & Golubov, J. (2004). Targeting sites for biological collections. In H. Pritchard & M. Way (Eds.), *Seed storage: Turning science into practice*. London, UK: Royal Botanic Gardens.
- Soberón, J., & Peterson, A. T. (2004). Biodiversity informatics: Managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359, 689–698. <https://doi.org/10.1098/rstb.2003.1439>
- Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2014). Retrieved from <http://www.openbugs.net/Manuals/Manual.html>
- Thomas, A., Best, N., Lunn, D., Arnold, R., & Spiegelhalter, D. (2004). Retrieved from <http://www.mrc-bsu.cam.ac.uk/software/bugs/>
- Vale, M., & Jenkins, C. N. (2012). Across-taxa incongruence in patterns of collecting bias. *Journal of Biogeography*, 39, 1743–1744. <https://doi.org/10.1111/j.1365-2699.2012.02750.x>
- van Eupen, M., Cormont, A., Kok, K., Simoes, M., Pereira, S., Kolb, M., & Ferraz, R. (2014). *Modelling land use change in Latin America*. Retrieved from <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/154803/1/ROBIN-D3.1.3-Stakeholder-methods-and-results-Embrapa-Luty-Maggie.pdf>
- Van Swaay, C. A. M., Nowicki, P., Settele, J., & Van Strien, A. J. (2008). Butterfly monitoring in Europe: Methods, applications and perspectives. *Biodiversity and Conservation*, 17, 3455–3469. <https://doi.org/10.1007/s10531-008-9491-4>
- Visconti, P., Pressey, R. L., Giorgini, D., Maiorano, L., Bakkenes, M., Boitani, L., ... Rondinini, C. (2011). Future hotspots of terrestrial mammal loss. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366, 2693–2702. <https://doi.org/10.1098/rstb.2011.0105>
- Wellcome, T. I., Kardynal, K. J., Franken, R. J., & Gillies, C. S. (2014). Off-road sampling reveals a different grassland bird community than roadside sampling: Implications for survey design and estimates to guide conservation. *Avian Conservation and Ecology*, 9, 4. <https://doi.org/10.5751/ACE-00624-090104>
- World Bank (2017). Retrieved from <http://data.worldbank.org/data-catalog/world-development-indicator>
- Yang, W., Ma, K., & Kreft, H. (2013). Geographical sampling bias in a large distributional database and its effects on species richness-environment models. *Journal of Biogeography*, 40, 1415–1426. <https://doi.org/10.1111/jbi.12108>
- Yang, W., Ma, K., & Kreft, H. (2014). Environmental and socio-economic factors shaping the geography of floristic collections in factors shaping the geography of. *Global Ecology and Biogeography*, 23, 1284–1292.
- Yesson, C., Brewer, P. W., Sutton, T., Caithness, N., Pahwa, J. S., Burgess, M., ... Culham, A. (2007). How global is the global biodiversity information facility? *PLoS ONE*, 2, e1124. <https://doi.org/10.1371/journal.pone.0001124>
- Zamora-Gutierrez, V., Pearson, R. G., Green, R. E., & Jones, K. E. (2018). Forecasting the combined effects of climate and land use change on Mexican bats. *Diversity and Distributions*, 24, 1–12. <https://doi.org/10.1111/ddi.12686>

How to cite this article: Zamora-Gutierrez V, Amano T, Jones KE. Spatial and taxonomic biases in bat records: Drivers and conservation implications in a megadiverse country. *Ecol Evol*. 2019;00:1–12. <https://doi.org/10.1002/ece3.5848>