Girardello, M., A. Chapman, R. Dennis, L. Kaila, P. A. V. Borges, and A. Santangeli. 2019. Gaps in butterfly inventory data: A global analysis. Biological Conservation 236:289-295.

Marco Girardello[a] Anna Chapman[b] Roger Dennis[c] Lauri Kaila[d] Paulo A.V. Borges[a] Andrea Santangeli[e]. 2019 Biological Conservation Vol. 236, 289-29

a

cE3c – Centre for Ecology, Evolution and Environmental Changes, Azorean Biodiversity Group and Universidade dos Açores, Faculty of Agriculture and Environment, Rua Capitão João d'Ávila, São Pedro, 9700-042 Angra do Heroísmo, Terceira, Açores, Portugal

b

Department of Biology, National and Kapodistrian University of Athens, Athens, Greece

c

Centre for Ecology, Environment and Conservation, Department of Biological and Medical Sciences, Oxford Brookes University, Oxford OX3 0BP, UK

d

Finnish Museum of Natural History, Zoology Unit, P.O. Box 17, FI-00014, University of Helsinki, Finland

e

The Helsinki Lab of Ornithology, Finnish Museum of Natural History, University of Helsinki, Finland

DOI: https://doi.org/10.1016/j.biocon.2019.05.053

Abstract

Species distribution data are crucial for assessing the conservation status of species (red listing, IUCN) and implementing international conservation targets, such as those set by the International Convention on Biological Diversity. Although there have been a number of efforts aimed at aggregating biodiversity data, information on the distribution of many taxa is still scanty (i.e. the Wallacean Shortfall). In this study, we use a large database, including over 19 million species occurrence records, to identify knowledge gaps in biodiversity inventories for butterfly records at a global level. Bayesian hierarchical spatial models were used to quantify the relationship between gaps in inventory completeness and the density of roads, protected areas and elevational range, the former variable being a proxy for accessibility, the latter two for attractiveness to recorders. Our results show that despite >100 years of butterfly sampling, knowledge of the distribution of butterflies is still limited in tropical areas. The results revealed that gaps in butterfly inventories are largely concentrated in areas of low elevational range, low density of protected areas and low road density. We conclude that the Wallacean Shortfall is a problem even for one of the best studied insect groups. In the light of these data limitations, we discuss prospects for filling gaps in butterfly inventories at the global scale within relatively short time frames. We argue that a combination of citizen science and quantitative tools may help to fill knowledge gaps and inform conservation decisions.

Key words: Aichi targets, biodiversity inventories; GBIF; knowledge-gaps; lepidoptera; Wallacean Shortfall

Introduction

The current biodiversity crisis calls for the urgent development of effective and efficient strategies to halt biodiversity loss and to preserve healthy ecosystems and the services they provide (Scholes et al., 2018). Conservation planning provides one of the most powerful frameworks for identifying priorities for conservation when resources are limited (Margules and Pressey, 2000). The main pillar of conservation planning is biodiversity data, so the availability and quality of these data are of paramount importance. The most common type of biodiversity data are occurrence records, i.e. observations of organisms, which include locations, times, and taxonomic identities. Such data are often associated with varying degrees of uncertainty, something which is important to consider when attempting to provide robust and informative outcomes from conservation planning exercises. Uncertainty in biodiversity data typically stems from limitations in the amount and quality of species level observations. Therefore, an important task for conservation scientists is to identify the gaps and shortfalls in biodiversity data (Hortal et al., 2015). This is particularly important when considering the large amount of digitally accessible information (DAI) now available, which has the potential to inform global conservation policy (Meyer et al., 2015).

Of all major animal groups, insects suffer the most from a paucity of observational data (Troudet et al., 2017). One of the many difficulties faced in the field of insect conservation is the Wallacean shortfall (Cardoso et al., 2011). This term describes the lack of knowledge regarding species distributions (Lomolino, 2004). Uncertainty in such distribution data can exist due to random errors introduced during data recording and storage, or due to spatio-temporal biases resulting from the choice of sampling sites or taxa (Chapman, 2005). Geographic bias occurs when sampling takes place more intensively or frequently in certain areas (Boakes et al., 2010). Researchers often concentrate their collection effort in areas of high accessibility, e.g. close to roads, or in areas known to be biodiversity hotspots, i.e. in species rich areas (Kadmon et al., 2004; Dennis and Thomas, 2000; Romo et al., 2006). These biases in observational data collection can affect the quality of results obtained when performing biogeographical and conservation planning studies (Yang et al., 2013). These include analyses of the large-scale determinants of species diversity and the use of occurrence records for predicting species distributions under different climate change scenarios.

For many major animal groups, geographic biases lead to unequal levels of data accumulation, resulting in spatial variation in inventory completeness. Regions with high accessibility, high levels of political stability and high levels of scientific funding generally have the most complete species inventories (Oliveira et al., 2016; Ballesteros-Mejia et al., 2013; Meyer et al., 2015). However, even in such regions, the accumulation of data is slow and the time required for achieving adequate levels of inventory completeness is long, as is the case for insect inventories (Fattorini, 2013; Lobo and Borges, 2010). It is therefore important to quantify temporal and spatial trends in data accumulation and to map current gaps in inventory completeness. This may encourage the redistribution of resources for compiling inventories towards regions that are currently poorly inventoried (Sánchez-Fernández et al., 2011; Amano and Sutherland, 2013).

Insect conservation has been hindered by difficulties in obtaining taxonomic, ecological and distribution data (New, 1997; Cardoso et al., 2011). A remarkable exception to this, within the megadiverse insect group, is the butterflies. Due to their relatively large size and aesthetic appeal, butterflies have traditionally been popular subjects of study for the general public as well as for scientists. This makes them important flagships for insect conservation (Barua et al., 2012). They also act as umbrella taxa for the protection of vulnerable habitat, and being easy to monitor, they are also valuable as indicators (New, 1997). Unlike many insect groups, butterflies are relatively easy to identify, even by non-specialists. Hence, parataxonomists and members of the public can be employed to monitor them without experiencing major difficulties (Schmiedel et al., 2016; Pollard and Yates, 1994). For these reasons, butterflies have become the focus of a number of high-profile citizen science projects and have contributed extensively to the current body of digitally accessible information (Gustafsson et al., 2015; Kocher and Williams, 2000).

A number of studies on inventory gaps have been performed for different groups of species and at various scales (Moreno and Halffter, 2001; Troudet et al., 2017). However, a global scale assessment of the gaps in butterfly inventories in time and space is still lacking. This may impair our ability to redistribute scarce resources and hinder progress towards setting conservation priorities for this important component of global biodiversity. Here, for the first time, we fill this gap in knowledge by quantifying the gaps in butterfly inventories at a global scale. Using a large amount of digitally accessible information spanning over 100 years and 205 countries, we first quantify temporal trends in data accumulation since the year 1900. Secondly, we map the gaps in butterfly inventories at the global scale. Thirdly, we test whether factors, such as area accessibility and appeal, have a role in the distribution of the completeness of inventories.

2. Methods

2.1. Species data

We downloaded all records for Lepidoptera from the Global Biodiversity Information Facility (GBIF) as of the 29th of January 2019 (https://www.gbif.org/occurrence/download/0028634-181108115102211). We filtered the GBIF search results to retain only records with coordinates and no documented geospatial issues. From this dataset, we extracted the data for butterflies, including all records for the families *Papilionidae, Hesperiidae, Hedylidae, Pieridae, Nymphalidae* and *Riodinidae*.

The dataset, including 19,228,465 records for 8652 species, was further filtered according to five additional criteria, leading to the overall exclusion of 56% of the initial records. We first excluded records of specimens with missing information on collection date, latitude and longitude. This resulted in the exclusion of 4% (769,138) of the records. Secondly, we excluded records with missing species names, which lead to the exclusion of 1% of the remaining records (210,071). Thirdly we excluded duplicate records (same species, coordinate and date combinations) and that further reduced the number of records by 59% (7,502,683 records). Next, we assessed the validity of the 8654 species names by submitting the species and subspecies names to the Global Name Resolver (http://gni.globalnames.org/). The Global Name Resolver provides a match score value per species ranging from zero to one, with one indicating a complete match between the input name checked and a valid taxon name in the core database, whereas a score of zero indicates no match. The Global Name Resolver uses exact and fuzzy matching criteria to check whether a string is a valid species name. Specifically, it checks whether: i) a name is spelt correctly, ii) the submitted name is the one currently in use, iii) there are any homonyms or synonyms and whether these include some lexical variants (https://resolver.globalnames.org/about). For this study, we selected only records for which the names of the species and subspecies matched with a score of > 0.9. This procedure led to the exclusion of a further 1062 records.

2.2. Assessing gaps in inventory completeness in space and time

Gaps in inventory completeness were assessed using smoothed species accumulation curves (SACs) (Yang et al., 2013). Species accumulation curves show the number of species obtained by successively sampling either individual organisms (individual-based accumulation curves) or samples (sample-based accumulation curves). SACs give the expected species richness for a certain number of records for a certain level of species richness. Accumulation curves tend towards a straight line in poorly sampled areas, and towards a high degree of curvature in better-sampled areas. The mean slope of the last 10% of SACs reflects the degree of curvilinearity and was used as a proxy for inventory completeness following (Yang et al., 2013). Shallow slopes (values close to 0) indicate saturation in the sampling (i.e. species accumulation has reached a plateau) and thus high levels of completeness, whereas steep slopes reflect low levels of completeness, i.e. large inventory gaps. Following Yang et al. (2013), we considered grid cells with slope values of ≤0.05 as well sampled, and those with slope values of >0.05 as under-sampled, i.e. with large inventory gaps. We also calculated another index of inventory gaps based on the Chao2 richness estimator (Chao, 1987). This is estimated using the ratio of the number of observed species from GBIF data to the number of species estimated through the Chao2 method. SACs were calculated with the method “exact” from the *specaccum* function in the “vegan” package in R (Oksanen et al., 2017). The slope was calculated through the *specslope* function which evaluates the derivative of the species accumulation curve for a given number of sample plots. Chao2 richness was estimated using the *chao2* function from the fossil package (Vavrek, 2011).

In order to explore spatial patterns in inventory gaps, SACs were calculated within defined sampling units referring to four different resolutions (i.e. cells of size: 110 km × 110 km, 220 km × 220 km, 440 km × 440 km, 880 km × 880 km) across the terrestrial areas of the globe. For each individual sampling unit, we considered the cumulative number of records and species collected from 1900 until 2018. In order to assess country-level trends in record accumulation, we fitted a series of Generalized Linear Models (GLM) to the total number of records per country, aggregated at the yearly level. The error structure associated with these models was assumed to be Poisson distributed with a log link function.

2.3. Correlates of inventory gaps

We used three layers to represent the degree of area accessibility and appeal. Road density was used as a proxy for the accessibility of collection areas. Road data were obtained from the digital chart of the world (Danko, 1992), and road density was calculated as the total length of roads (in km) divided by the area of each grid cell. Area appeal was represented by the density of protected areas and the presence of mountainous areas (expressed as the elevational range for each grid cell). We chose these two variables to represent attractiveness, as they would draw the attention of collectors, being considered “pristine areas”, as well as being strongholds of rare species and areas of scenic beauty. Density of protected areas was calculated using the WDPA database (http://www.wdpa.org/, and included IUCN protected areas in categories I to VI). Elevation range was derived from the GMTED2010 digital elevation model (https://topotools.cr.usgs.gov/gmted_viewer/viewer.htm). Spatial linear models were fitted to the response variable (the score of inventory completeness) using the aforementioned variables as predictors. The models were fitted within a Bayesian framework using integrated nested Laplace approximations (INLA). INLA were chosen as they allow efficient estimation of regression parameters within a Bayesian framework, without the need to employ computationally intensive Markov Chain Monte Carlo algorithms. A spatially correlated random intercept term was incorporated into the models. The random intercept is assumed to be a Gaussian Markov Random Field (GMRF) with mean 0 and covariance matrix $\sum$. Because of the difficulty in estimating a covariance matrix for large datasets, an approach based on Continuous Domain Stochastic Partial Differential Equations (SPDE) is used to calculate the covariance matrix $\sum$ for the GMRF (Lindgren et al., 2011). This is a computationally efficient approach for large datasets, as it avoids the computation of a full covariance matrix for the GMRF. Default priors were assigned for all fixed-effect parameters, as recommended by Held et al. (2010), which are approximations of non-informative priors designed to have little influence on the posterior distribution.

3. Results

3.1. Trends in record accumulation

A substantial increase in the amount of occurrence records occurred from 1990 onwards, with almost a three-fold increase in record accumulation compared to the period before 1990 (Fig. 1). With the exception of a few countries, trends in record accumulation for the period between 1900 and 2018 were positive across much of the world. However, this was not the case for central Africa (Fig. 2, Fig. S4). Yet a comparison of trends over shorter periods of time showed that trends were mostly idiosyncratic and did not display any defined spatial patterns. With the exception of Kazakhstan and the Ukraine, trends from 1950 until 1990 were positive over much of the Palearctic and Nearctic regions. Conversely, many countries in the southern hemisphere displayed a decreasing trend for the same period. Despite the massive increase in the number of records during 1990–2018, this trend was not homogeneous throughout the world. For this recent period, butterfly record accumulation appears to decrease in several countries in the northern hemisphere as well as in the tropics (Fig. 2, Fig. S4).

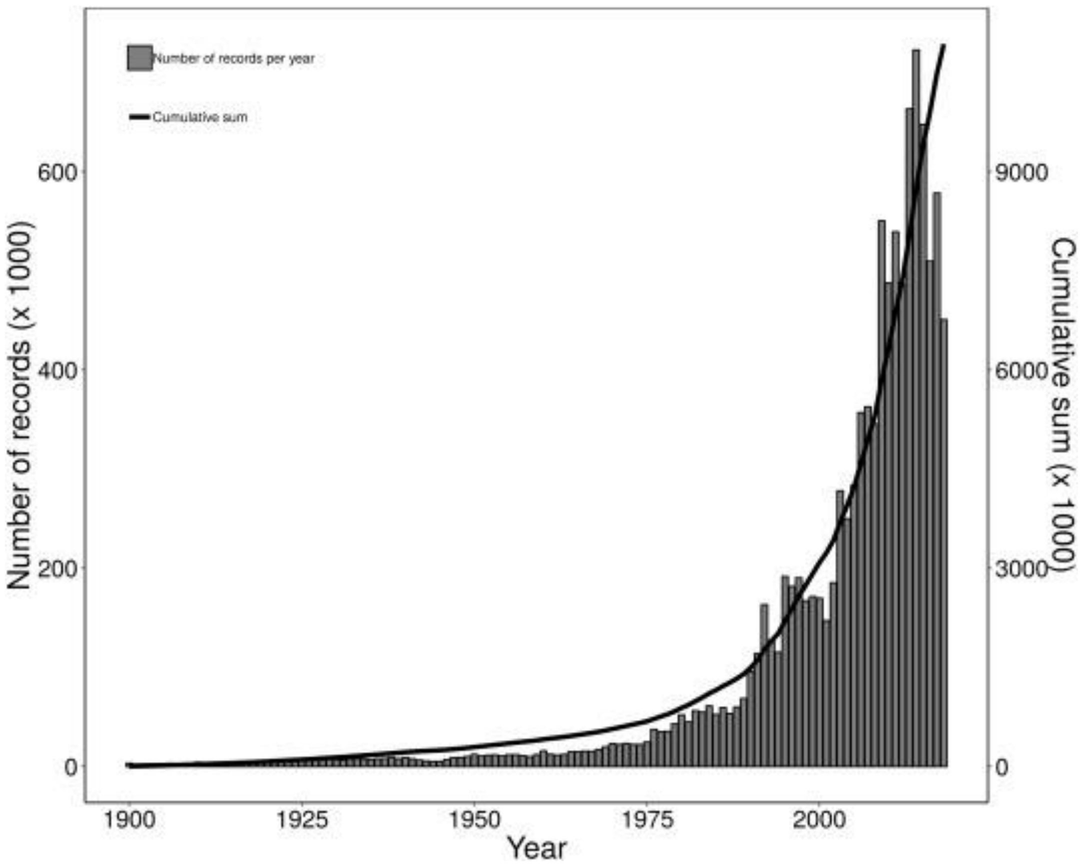

Fig. 1. Global accumulation in the number of butterfly observations/specimens from the year 1900 until 2018. Bars represent the total number of occurrence records for a given year. The black line represents the cumulative sum of the number of records.

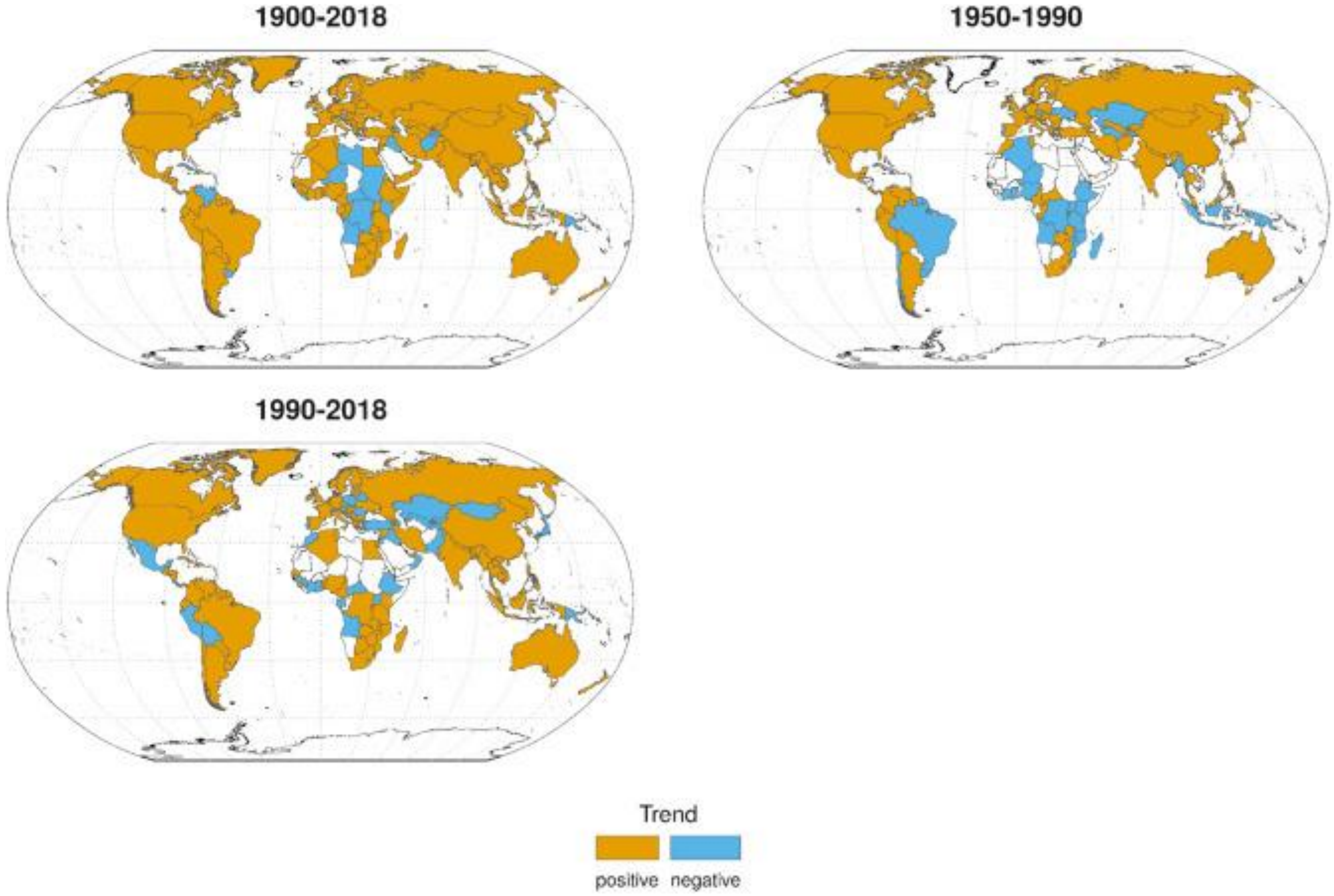

Fig. 2. Global accumulation in the number of butterfly observations/specimens from the year 1900 until 2018. Trends in data accumulation for butterflies calculated at the country level. Trends were calculated for different periods using Generalized Linear Models (GLM) fitted to the number of occurrence records aggregated by different year/country combinations. Countries shown in white have been omitted from analyses due the absence of records for a given period.

3.2. Spatial distribution of inventory completeness

We found clear spatial biases in the dataset used, with striking regional differences. Inventory completeness was highest in Europe, North America, coastal regions of Australia, and Southern Africa, whereas most inventory gaps were spread across large regions of South America, Saharan Africa, West Africa, Asia and Oceania (excluding Australia and New Zealand). Overall, there was a high level of spatial clustering within areas with high levels of inventory completeness in temperate and tropical areas alike. Coarsening the grain of the analysis would increase completeness (Fig. 3). A map based ratio of the number of observed species from GBIF data (see Fig. S2) to the number of species estimated using the Chao2 estimator (Fig. S3) showed very similar results, strongly correlated for all grain sizes (mean rho = 0.85).

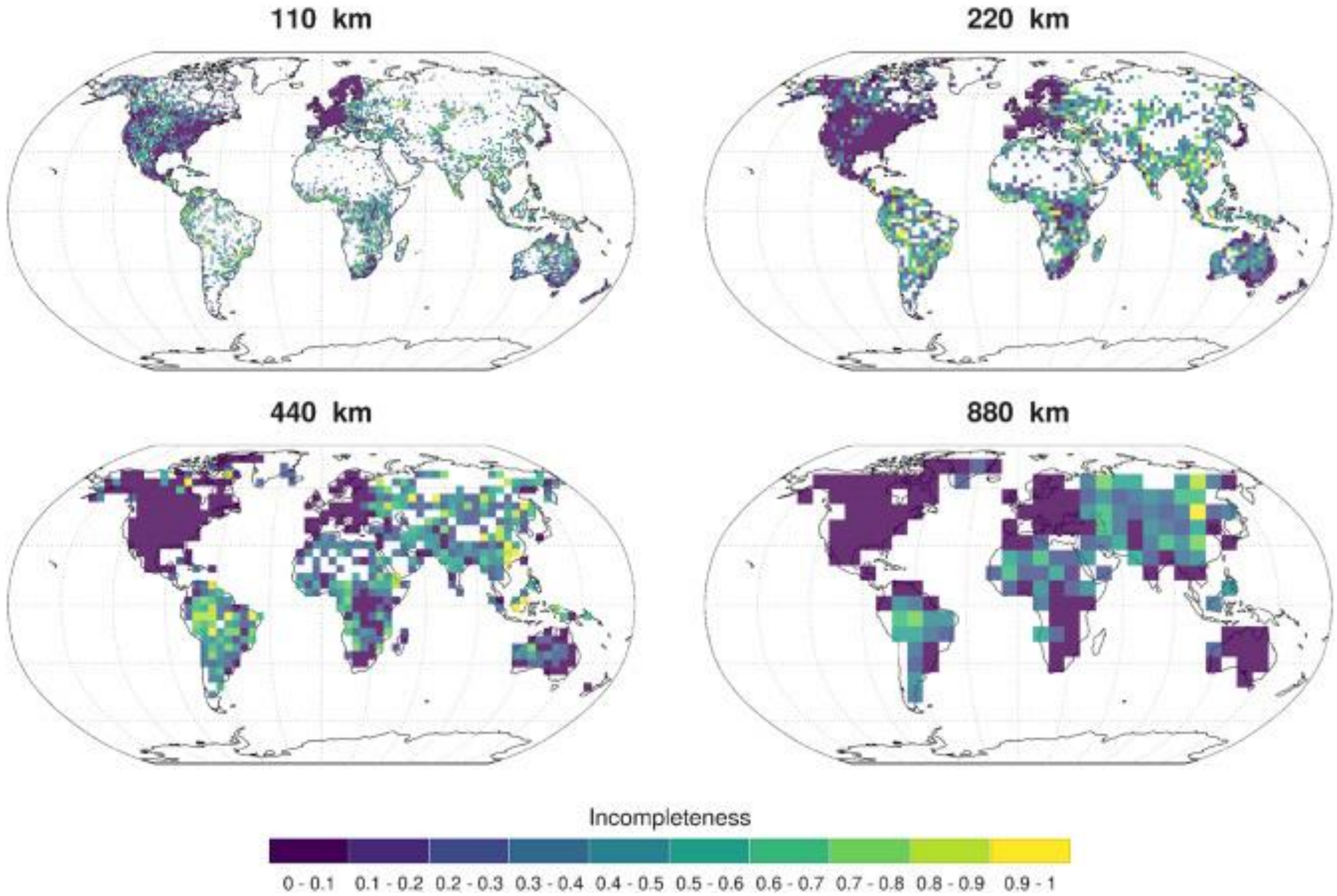

Fig. 3. Global gaps in butterfly inventories (i.e. incompleteness). Inventory incompleteness was calculated as the slope of the last 10% of species accumulation curves for grid cells at four different resolutions. Slope values were rescaled, with values close to 0 indicating highly complete inventories and values close to 1 indicating highly incomplete inventories, i.e. large inventory gaps. Blank areas indicate squares that did not have any records.

3.3. Correlates of inventory gaps

Model selection based on the Deviance Information Criterion (DIC) revealed important differences between the best ranking models fitted to data available at different grain sizes (Table 1 and Table S1). Road density was retained in the best model fitted to the data at a resolution of 110 km. Altitudinal range was retained in the best models fitted to the dataset at a resolution of 440 km and 880 km. Lastly, density of protected areas was retained only in the best model fitted to the dataset at a resolution of 880 km. Overall, gaps in butterfly inventories were largely concentrated in areas of low elevational range, low density of protected areas and low road density (Table 1). No predictors were retained for the best model developed using data at a resolution of 220 km. A graphical inspection of the posterior distributions for the parameters of the best ranking models showed that these did not overlap with 0 (Fig. S5).

Table 1. Parameter estimates for the best linear spatial regression models based on summaries of the marginal posterior distributions of the predictors. Models were fitted to the incompleteness scores, calculated as the slope of the last 10% of species accumulation curves for grid cells at four different resolutions. High values for the slope, the response variable, indicate larger gaps in butterfly inventory. Predictors were standardized. Model selection performed using the Deviance Information Criterion (DIC). Numbers between square brackets indicate 95% Bayesian credibility intervals.

Predictor	**Resolution**			
	100 km x 100 km	**220 km x 220 km**	**440 km x 440 km**	**880 km x 880 km**
Protected Area Density	-	-	-	-0.015[-0.0372;0.0072]
Elevational Range	-	-	-0.0128[-0.029;0.004]	-0.0133[-0.013;0.013]
Road Density	-0.009[-0.016;0.002]	-	-	

4. Discussion

We show that globally butterfly data accumulation has been progressively increasing, particularly since 1990. Interestingly, during this recent period, data accumulation has broadly slowed in many countries of the Northern Hemisphere, whereas the opposite is generally the case in the Global South. We also show that while inventory completeness is relatively good for the Global North, large gaps exist in the Southern Hemisphere, particularly in Asia and South America. Finally, our results confirm the widely acknowledged pattern that the best inventoried areas are those that have the highest appeal among researchers and are more accessible.

Despite the pervasive spatial bias in inventory completeness, the positive general trend in data accumulation of butterfly occurrence, particularly over recent decades, is encouraging. It is likely to highlight a broadly increasing interest in butterfly inventories that is translated into efforts to collect more data, or to make these data more available on GBIF. This trend aligns with Aichi Target 19 of the Convention on Biological Diversity, which states that by the year 2020, knowledge of biodiversity should be improved, shared, transferred and put into practice (Secretariat of the Convention on Biological Diversity, 2014). On the other hand, the large gaps in butterfly inventory data across several regions of the world, particularly in the Global South, suggest that a great effort must be made to accumulate more data in these regions. Only then will we be able to achieve a thorough understanding of butterfly occurrence and distribution. As such, our findings provide further support to the assertion that the ambitious Aichi target 19 will not easily and feasibly be met within the set timeframe (Secretariat of the Convention on Biological Diversity, 2014).

However, we also found some signals over the past century, of a potential redistribution, or shift in effort regarding butterfly inventories towards poorly surveyed areas. We show that trends in butterfly data accumulation have recently become more positive in several countries of the Global South (Fig. 2). Increased attempts to address the so-called Wallacean shortfall (i.e. lack of distribution data) in butterfly conservation, may be partly responsible for this increased inventory effort in the Global South. Despite these positive trends, we are still a long way from addressing the Wallacean shortfall and strategies need to be put in place that will mitigate this limitation.

We show that the southern regions of the world are associated with very large gaps in butterfly inventory, a pattern that is broadly common to most other taxa, such as plants and vertebrates, including freshwater fishes (Meyer et al., 2016; Meyer et al., 2015; Pelayo-Villamil et al., 2018). Recent scientific interest in highlighting gaps in knowledge (Amano et al., 2016; Buechley et al., 2019; Troudet et al., 2017) coupled with the designation of global targets for improving biodiversity knowledge and its transfer (Aichi Target 19) may be in part responsible for the spatial variation in butterfly data accumulation trends found here. Similarly, the increased mobility of researchers and citizens alike over the past decades, coupled with improved geopolitics in previously unsafe regions (e.g. Central Africa) may have also contributed to the generally positive trend in butterfly data accumulation in the global South.

The spatial bias for greater inventory completeness in mountainous areas mirrors patterns previously reported for other taxa (Meyer et al., 2015; Yang et al., 2014). A strong incentive for data collection is thought to be regional or species ‘appeal’. For example, researchers appear to show a preference for reserves, mountains or other areas of high total, rare and range-restricted species richness. Surprisingly, we found that accessibility (i.e. road density) was only important at the finest resolution (110 km), which has been previously reported for other taxa (Meyer et al., 2015).

It is important to note that the results we report are only relevant with regard to the GBIF data we used. Such data may suffer from spatial and temporal biases in the way they are made available. For example, for some regions of the world, much inventory data, especially historical data, may be available, but they are not stored in electronic databases, or specimens are still waiting for identification due to a lack of resources and/or capacity (Anderson et al., 2016; García-Roselló et al., 2015). However, GBIF is arguably the most complete and standardized platform for reporting and downloading global butterfly data. As such, it is also one of the most valuable databases for conservation planning. Its potential use for identifying priority areas for conservation thus further emphasizes the relevance of this study's results in addressing its limitations in spatial coverage and in promoting a better distribution of inventory effort to overcome spatial biases.

The gross paucity of butterfly inventory data reported here, particularly from Asia and South America, raises deep concerns. These regions are under high anthropogenic pressure from land-use change and intensification (Asselen and Verburg, 2013) and they have been recently identified as largely lacking conservation capacity (i.e. a sufficient number of local conservationists with adequate skills and expertise) (Elliott et al., 2018). Therefore, capacity development, relative to both biodiversity monitoring and conservation (i.e. boosting the skills and expertise of local conservationists and biodiversity observers) in these poorly monitored regions will be paramount in order for them to meet the Aichi Biodiversity Targets of the CBD (Vanhove et al., 2017). Failure to fill the capacity development gap in the Global South will hinder implementation of timely and effective evidence-based conservation, leading to further biodiversity declines.

While most efforts in improving data quality and quantity for vertebrate species have been put in place during the past decades, butterflies, as is the case for most other invertebrate groups, are lagging far behind (Troudet et al., 2017). Given their importance as indicators of climate change (Oliver et al., 2015; Devictor et al., 2012; Chen et al., 2011), there is an impelling need to increase the amount of information on butterfly species occurrence, particularly in the tropics. A roadmap for achieving better completeness of distribution data should firstly include (as stated in Aichi Target 20) the mobilization of international funds to perform standardized inventories in poorly studied regions. This would involve incorporating local scientists and NGOs in the process, so creating local capacity and overcoming the increasing difficulty of issuing research permits to qualified personnel to enable essential fieldwork. Secondly, it should include the digitization of inventories and collections that exist outside standardized databases such as GBIF. Thirdly, the availability of raw species distribution data in publications must be made mandatory, a process already being implemented in many scientific journals, which will facilitate data acquisition for GBIF. Fourthly, it will include the promotion of citizen science projects for lesser known organisms in poorly surveyed regions. Ultimately, the emergence and spread of technology, with ad-hoc, accessible and easily available mobile phone applications for identifying and reporting biodiversity observations from almost anywhere in the world is highly promising with regard to filling current gaps in biodiversity inventories (Pimm et al., 2015). However, achieving these goals for all taxa within a short time frame is not realistic, and priorities need to be put in place. One possibility will be the mobilization of the butterfly scientific community to perform standardized inventories in tropical areas, following the example of recent efforts, such as the Global Island Monitoring Scheme (GIMS) of forest biota across islands (Borges et al., 2018). In addition to this, more effort is needed to understand how to make the most use of incomplete information. This would include the development of appropriate analytical methods for analyzing sparse data and the use of flexible modelling tools for overcoming data limitations (Isaac et al., 2014).

References

Anderson, R. P., Araújo, M. B., Guisan, A., Lobo, J. M., Martínez-Meyer, E., Peterson, A. T., Soberón, J. (2016). Are species occurrence data in global online repositories fit for modeling species distributions? The case of the Global Biodiversity Information Facility (GBIF). Final Report of the Task Group on GBIF Data Fitness for Use in Distribution Modelling. Global Biodiversity Information Facility (GBIF).

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. Proc. R. Soc. B 280, 20122649.

Asselen, S., Verburg, P.H., 2013. Land cover change or land-use intensification: simulating land system change with a global-scale land change model. Glob. Chang. Biol. 19, 3648–3667.

Ballesteros-Mejia, L., Kitching, I.J., Jetz, W., Nagel, P., Beck, J., 2013. Mapping the biodiversity of tropical insects: species richness and inventory completeness of African sphingid moths. Glob. Ecol. Biogeogr. 22, 586–595. https://doi.org/10.1111/geb.12039

Barua, M., Gurdak, D.J., Ahmed, R.A., Tamuly, J., 2012. Selecting flagships for invertebrate conservation. Biodivers. Conserv. 21, 1457–1476.

Boakes, E.H., McGowan, P.J.K., Fuller, R.A., Chang-qing, D., Clark, N.E., O'Connor, K., Mace, G.M., 2010. Distorted Views of Biodiversity: Spatial and Temporal Bias in Species Occurrence Data. PLOS Biol. 8, 1–11. https://doi.org/10.1371/journal.pbio.1000385

Borges, P.A. V, Cardoso, P., Kreft, H., Whittaker, R.J., Fattorini, S., Emerson, B.C., Gil, A., Gillespie, R.G., Matthews, T.J., Santos, A.M.C., Steinbauer, M.J., Thébaud, C., Ah-Peng, C., Amorim, I.R., Aranda, S.C., Arroz, A.M., Azevedo, J.M.N., Boieiro, M., Borda-de-Água, L., Carvalho, J.C., Elias, R.B., Fernández-Palacios, J.M., Florencio, M., González-Mancebo, J.M., Heaney, L.R., Hortal, J., Kueffer, C., Lequette, B., Martín-Esquivel, J.L., López, H., Lamelas-López, L., Marcelino, J., Nunes, R., Oromí, P., Patiño, J., Pérez, A.J., Rego, C., Ribeiro, S.P., Rigal, F., Rodrigues, P., Rominger, A.J., Santos-Reis, M., Schaefer, H., Sérgio, C., Serrano, A.R.M., Sim-Sim, M., Stephenson, P.J., Soares, A.O., Strasberg, D., Vanderporten, A., Vieira, V., Gabriel, R., 2018. Global Island Monitoring Scheme (GIMS): a proposal for the long-term coordinated survey and monitoring of native island forest biota. Biodivers. Conserv.

27, 2567–2586. https://doi.org/10.1007/s10531-018-1553-7

Cardoso, P., Erwin, T.L., Borges, P.A. V, New, T.R., 2011. The seven impediments in invertebrate conservation and how to overcome them. Biol. Conserv. 144, 2647–2655.

Buechley, E.R., Santangeli, A., Girardello, M., Neate-Clegg, M.H.C., Oleyar, D., McClure, C.J.W., Şekercioğlu, Ç.H., 2019. Global raptor research and conservation priorities: Tropical raptors fall prey to knowledge gaps. Diversity and Distributions, *in press*.

Chapman, A.D., 2005. Principles and methods of data cleaning: primary species and species-occurrence data. Global Biodiversity Information Facility.

Chao, A. 1987. Estimating the population size for capture-recapture data with unequal catchability. Biometrics 43:783–791.

Chen, I.-C., Hill, J.K., Ohlemüller, R., Roy, D.B., Thomas, C.D., 2011. Rapid range shifts of species associated with high levels of climate warming. Science (80-.). 333, 1024–1026.

Danko, D.M., 1992. The digital chart of the world project. Photogramm. Eng. Remote Sensing 58, 1125–1128.

Dennis, R.L.H., Thomas, C.D., 2000. Bias in Butterfly Distribution Maps: The Influence of Hot Spots and Recorder's Home Range. J. Insect Conserv. 4, 73–77. https://doi.org/10.1023/A:1009690919835

Devictor, V., van Swaay, C., Brereton, T., Brotons, L., Chamberlain, D., Heliölä, J., Herrando, S., Julliard, R., Kuussaari, M., Lindström, Å., Reif, J., Roy, D.B., Schweiger, O., Settele, J., Stefanescu, C., Van Strien, A., Van Turnhout, C., Vermouzek, Z., WallisDeVries, M., Wynhoff, I., Jiguet, F., 2012. Differences in the climatic debts of birds and butterflies at a continental scale. Nat. Clim. Chang. 2, 121.

Diversity, S. of the C. on B., 2014. No Title. Glob. Biodivers. Outlook 4. Page 155, Montréal.

Elliott, L., Ryan, M., Wyborn, C., 2018. Global patterns in conservation capacity development. Biol. Conserv. 221, 261–269.

Fattorini, S., 2013. Regional Insect Inventories Require Long Time, Extensive Spatial Sampling and Good Will. PLoS One 8, 1–9. https://doi.org/10.1371/journal.pone.0062118

García-Roselló, E., Guisande, C., Manjarrés-Hernández, A., González-Dacosta, J., Heine, J., Pelayo-Villamil, P., ... & Lobo, J. M. (2015). Can we derive macroecological patterns from primary Global Biodiversity Information Facility data?. Global Ecology and Biogeography, 24(3), 335-347.

Gustafsson, K.M., Agrawal, A.A., Lewenstein, B. V, Wolf, S.A., 2015. The Monarch Butterfly through Time and Space: The Social Construction of an Icon. Bioscience 65, 612–622. https://doi.org/10.1093/biosci/biv045

Held, L., Schrödle, B., Rue, H., 2010. Posterior and cross-validatory predictive checks: a comparison of MCMC and INLA. Stat. Model. Regres. Struct. 91–110.

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. Annu. Rev. Ecol. Evol. Syst. 46, 523–549. https://doi.org/10.1146/annurev-ecolsys-112414-054400

Isaac, N.J.B., van Strien, A.J., August, T.A., de Zeeuw, M.P., Roy, D.B., 2014. Statistics for citizen science: extracting signals of change from noisy ecological data. Methods Ecol. Evol. 5, 1052–1060. https://doi.org/10.1111/2041-210X.12254

Kadmon, R., Farber, O., Danin, A., n.d. Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. Ecol. Appl. 14, 401–413. https://doi.org/10.1890/02-5364

Kocher, S.D., Williams, E.H., 2000. The diversity and abundance of North American butterflies vary with habitat disturbance and geography. J. Biogeogr. 27, 785–794.

Lindgren, F., Rue, H., Lindström, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J. R. Stat. Soc. Ser. B (Statistical Methodol. 73, 423–498.

Lobo, J.M., Borges, P.A. V, 2010. The provisional status of arthropod inventories in the Macaronesian islands. Terr. arthropods Macaronesia--Biodiversity, Ecol. Evol. 33–47.

Lomolino, M. V, 2004. Conservation biogeography. Front. Biogeogr. new Dir. Geogr. Nat. 293–296.

Margules, C.R., Pressey, R.L., 2000. Systematic conservation planning. Nature 405, 243–253. https://doi.org/10.1038/35012251

Meyer, C., Kreft, H., Guralnick, R., Jetz, W., 2015. Global priorities for an effective information basis of biodiversity distributions. Nat. Commun. 6, 8221. https://doi.org/10.1038/ncomms9221

Meyer, C., Weigelt, P., Kreft, H., 2016. Multidimensional biases, gaps and uncertainties in global plant occurrence information. Ecol. Lett. 19, 992–1006. https://doi.org/10.1111/ele.12624

Mora, C., Tittensor, D.P., Myers, R.A., 2008. The completeness of taxonomic inventories for describing the global diversity and distribution of marine fishes. Proc. R. Soc. London B Biol. Sci. 275, 149–155.

Moreno, C.E., Halffter, G., 2001. On the measure of sampling effort used in species accumulation curves. J. Appl. Ecol. 38, 487–490.

New, T.R., 1997. Are Lepidoptera an effective ‘umbrella group ‘for biodiversity conservation? J. Insect Conserv. 1, 5–12.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O’Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., 2017. vegan: Community Ecology Package.

Oliveira, U., Paglia, A.P., Brescovit, A.D., de Carvalho, C.J.B., Silva, D.P., Rezende, D.T., Leite, F.S.F., Batista, J.A.N., Barbosa, J.P.P.P., Stehmann, J.R., Ascher, J.S., de Vasconcelos, M.F., De Marco, P., Löwenberg-Neto, P., Dias, P.G., Ferro, V.G., Santos, A.J., 2016. The strong influence of collection bias on biodiversity knowledge shortfalls of Brazilian terrestrial biodiversity. Divers. Distrib. 22, 1232–1244. https://doi.org/10.1111/ddi.12489

Oliver, T.H., Marshall, H.H., Morecroft, M.D., Brereton, T., Prudhomme, C., Huntingford, C., 2015. Interacting effects of climate change and habitat fragmentation on drought-sensitive butterflies. Nat. Clim. Chang. 5, 941.

Pelayo-Villamil, P., Guisande, C., Manjarrés-Hernández, A., Jiménez, L.F., Granado-Lorencio, C., García-Roselló, E., González-Dacosta, J., Heine, J., González-Vilas, L., Lobo, J.M.J.B., 2018. Completeness of national freshwater fish species inventories around the world. Biodiversity and Conservation. 27, 3807-3817.

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 Ecol. Evol. 30, 685–696. https://doi.org/https://doi.org/10.1016/j.tree.2015.08.008

Pollard, E., Yates, T.J., 1994. Monitoring butterflies for ecology and conservation: the British butterfly monitoring scheme. Springer Science & Business Media.

Romo, H., García-Barros, E., Lobo, J.M., 2006. Identifying recorder-induced geographic bias in an Iberian butterfly database. Ecography (Cop.). 29, 873–885. https://doi.org/10.1111/j.2006.0906-7590.04680.x

Sánchez-Fernández, D., Lobo, J.M., Hernández-Manrique, O.L., 2011. Species distribution models that do not incorporate global data misrepresent potential distributions: A case study using Iberian diving beetles. Divers. Distrib. 17, 163–171. https://doi.org/10.1111/j.1472-4642.2010.00716.x

Schmiedel, U., Araya, Y., Bortolotto, M.I., Boeckenhoff, L., Hallwachs, W., Janzen, D., Kolipaka, S.S., Novotny, V., Palm, M., Parfondry, M., Smanis, A., Toko, P., 2016. Contributions of paraecologists and parataxonomists to research, conservation, and social development. Conserv. Biol. 30, 506–519.

Scholes, R., Montanarella, L., Brainich, A., Barger, N., ten Brink, B., Cantele, M., Erasmus, B., Fisher, J., Gardner, T., Holland, T.G., others, 2018. Summary for Policymakers of the Thematic assessment Report on Land Degradation and Restoration of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. IPBES Secr. Bonn, Ger. 1–31.

Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R., Legendre, F., 2017. Taxonomic bias in biodiversity data and societal preferences. Sci. Rep. 7, 9132. https://doi.org/10.1038/s41598-017-09084-6

Vanhove, M.P.M., Rochette, A.-J., de Bisthoven, L.J., 2017. Joining science and policy in capacity development for monitoring progress towards the Aichi Biodiversity Targets in the global South. Ecol. Indic. 73, 694–697.

Vavrek, M. J. (2011). Fossil: palaeoecological and palaeogeographical analysis tools. Palaeontologia Electronica, 14(1), 16.

Yang, W., Ma, K., Kreft, H., 2014. Environmental and socio-economic factors shaping the geography of floristic collections in China. Glob. Ecol. Biogeogr. 23, 1284–1292. https://doi.org/10.1111/geb.12225

Yang, W., Ma, K., Kreft, H., 2013. Geographical sampling bias in a large distributional database and its effects on species richness–environment models. J. Biogeogr. 40, 1415–1426. https://doi.org/10.1111/jbi.12108