

# Deviating environmental niche estimations from GBIF occurrence data and IUCN range maps

Elsa Östlund

elsa.ostlund@unil.ch

Department of Ecology and Evolution, University de Lausanne, Switzerland

1st step project, directed by Antoine Guisan, supervised by Mathieu Chevalier.  
Master in Behaviour, Evolution and Conservation, University of Lausanne, 2020

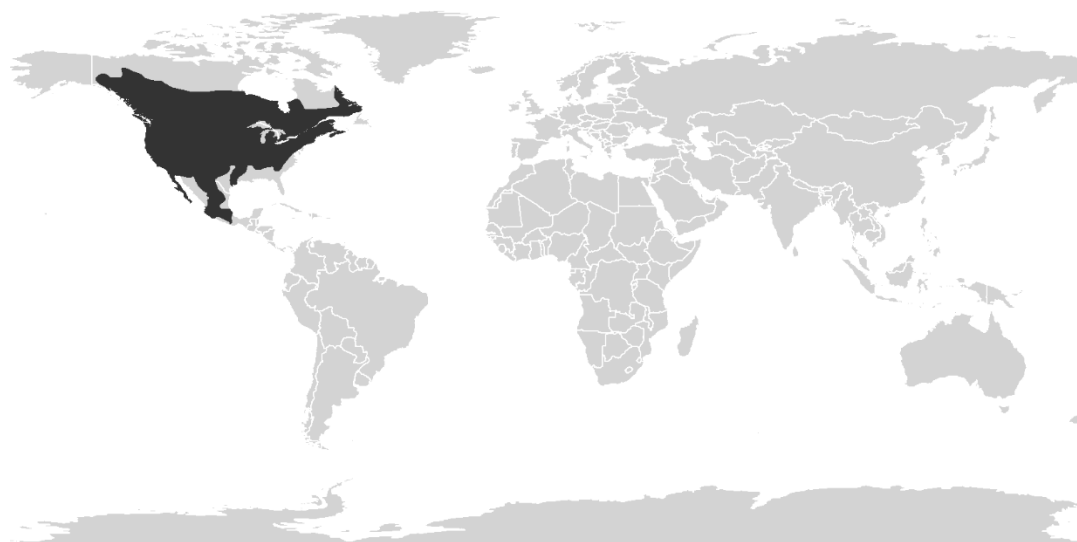
## Abstract

Many macroecological studies are relating species distributions to environmental data to gain knowledge about the factors affecting the distribution of species across large spatial ranges. This knowledge can be used to quantify the environmental niche of species. Species distributions can be obtained through different data sources but there are two sources that have global coverage: range maps provided by the International Union for Conservation of Nature and georeferenced occurrence records provided by the Global Biodiversity Information Facility. Both databases are frequently used for their high taxonomic, temporal, and spatial coverage, although both are affected by well-known biases. Therefore, knowing how the choice of data influences the tendency of the estimations is very important. A previous study concluded that environmental data estimates for rodent species derived from the two datatypes are highly correlated. Here I show that a multivariate approach reveals the strong influence of the data choice, when estimating environmental niches of rodents. Similar discrepancies could also exist for other taxa which would be interesting to investigate further. To accurately produce environmental niche models and niche estimations is of great importance for many fields in ecology, but especially for conservation research and planning.

## Introduction

Many macroecological studies are relating species distributions to environmental data to gain knowledge about the factors affecting the distribution of species across large spatial ranges. This knowledge can then be used to estimate current species distributions (Soberón, 2007), reveal species specialisation (Fernanda Bonetti & Wiens, 2014) and quantify the realized niche of a species (Vetaas, 2002), which is the set of suitable and accessible environmental conditions constrained by biotic interactions. It can also be used to test biogeographical theories like the “abundant centre” hypothesis (Sagarin & Gaines, 2002) or investigate how species’ ranges could shift between different time periods, for example due to climate change (Guisan & Thuiller, 2005; Pearson & Dawson, 2003; Thuiller, 2004).

The distribution of species can be obtained through different data sources but there are two sources that have global coverage: 1) Range maps from the International Union for Conservation of Nature, created by experts in the field. Experts with specific knowledge about a specie’s ecology and distribution patterns can identify areas where there has been records of the species and where the habitat is suitable. The IUCN has more than >96,600 species covered for spatial data (IUCN Red List, 2020). The maps are based on simplified polygons defined around species locations (Figure 1) according to knowledge about the species ecological constraints (Herkt, Skidmore, & Fahr, 2017).



**Figure 1:** Example of range map provided by the International Union for Conservation of Nature. Black polygon represents the estimated distribution of *Peromyscus maniculatus*.

2) The second data source is georeferenced species occurrence records from the Global Biodiversity Information Facility (GBIF; Figure 2). The occurrences are derived from observations in the field and have worldwide records of more than one million species that are either registered by professionals in the field or by private individuals. GBIF is based on extensive observation data collected mainly through citizen science projects for a diverse range of taxonomic groups and across large geographical and temporal scales. This database is increasingly valuable for conservation and research because of the wide coverage that is made possible through volunteer participation (Schmeller et al., 2009). The GBIF occurrence database is frequently used to derive exact environmental variables at the

species location which can be used to interpret the macroecological patterns of the species (Savage & Vellend, 2015).



**Figure 2:** Example of georeferenced occurrence records provided by the Global Biodiversity Information Facility. Black points represent the estimated distribution of *Peromyscus maniculatus*.

Both databases are frequently used to derive information about species realized niches by cross-referencing occurrence records with environmental databases (Martínez-Méndez, Aguirre-Planter, Eguiarte, & Jaramillo-Correa, 2016; Warren, Price, Graham, Forstenhaeusler, & VanDerWal, 2018). This information can be used to train species distribution models (SDM) (R. Boria & Blois, 2018; Gaikwad, Wilson, & Ranganathan, 2011), which is a way to model a species niche. It is also used to extract specific information about niche positions or niche breadth (Curtis & Bradley, 2016) and thereby to estimate how specialized a species is.

Although the two data sources provide accessibility to big datasets of importance to many ecological studies (Guisan & Thuiller, 2005), both datasets are affected by established biases (Beck, Böller, Erhardt, & Schwanghart, 2014; Herkt et al., 2017). The GBIF database has many biases that arise from the citizen-science based data collection. Data is collected by volunteers and is not done randomly. Sampling sites are often chosen because of their accessibility and attractiveness (Romo, García-Barros, & Lobo, 2006) but also because of their proximity to the home base of the recorder and if the sites are considered species hot-spots (Dennis & Thomas, 2000). The influence of the biases tends to vary depending on the taxonomic group, since volunteers usually focus on only a few groups (Snäll, Kindvall, Nilsson, & Pärt, 2011). Often, they also do not record every species they see, either because of lack of interest in very common species (Snäll et al., 2011), or because of lack of identification skills (Dickinson, Zuckerberg, & Bonter, 2010). This means that some areas are surveyed a lot, while others receive little or no survey effort, resulting in records that have an uneven cover of the whole distribution of the species. GBIF is based on presence-only data which gives rise to another bias. When designing distribution models from these observations, unintentional weight is given to the absence of records in a site. Without absence data, the species distribution cannot be distinguished from the distribution of the volunteers (Isaac, van Strien, August, de Zeeuw, & Roy, 2014). These biases are sometimes

accounted for by thinning the data (R. A. Boria, Olson, Goodman, & Anderson, 2014; R. Boria & Blois, 2018; Fourcade, Engler, Rödder, & Secondi, 2014). In practice this can be done by overlaying a raster layer with certain resolution on the occurrence record and only keep one occurrence per pixel.

The IUCN range maps also endure biases, mostly arriving from the design of the polygons. The polygons are drawn in a rough manner and the range boundaries usually ignore the fact that species may not be found everywhere within the polygon (e.g. mountain tops or lakes) (Brown, Stevens, & Kaufman, 1996). Since the IUCN range maps were created to guide efforts in conservation, they have been drawn in a conservative manner (Herkt et al., 2017). Experts were asked to only use range predictions with presumed suitable habitats, in which the species is known to occur, and not include habitats where the species is suspected to occur (Herkt et al., 2017). They were also asked to disregard localities that were far from the other confirmed presence areas (Schipper et al., 2008). These biases are important to keep in mind since they lead to systematic underestimations of the species' geographical range (Feng, Castro, McBee, & Papes, 2017; Gaston & Fuller, 2009). In poorly surveyed regions, that often harbour high species richness, these generalisations generate even bigger biases (Ficetola et al., 2014). Therefore, to use this kind of range maps for macroecological research have been advised against (Herkt et al., 2017), but no consensus have been reached on which source of distribution data is best to use for these questions (Alhajeri & Fourcade, 2019; Beck et al., 2014; Fourcade, 2016; Tiago, Pereira, & Capinha, 2017).

The reason why these databases are still used, although their biases, is that they have a very high taxonomic, temporal, and spatial coverage. For species distribution and environmental niche modelling these databases are very useful, but the biases must be considered when choosing the data source. Knowing how the data collection influences the tendency of the data is thus very important. Recently, Alhajeri & Fourcade (2019) tested whether the environmental conditions used by species across their geographical range are consistent between the two databases, IUCN and GBIF. They chose to approach this comparison without thinning the GBIF data and using a univariate method, which means that they only looked at one environmental variable at a time. They showed a high correlation between the environmental data estimates derived from the two datatypes and concluded that the two databases would provide similar results and can be used interchangeably.

The objective of this study is to investigate if there is high uniformity in the environmental conditions that can be extracted from the species' occurrences/ranges in the two different databases. First, the same univariate approach as Alhajeri & Fourcade (2019) will be assessed and then a multivariate method that is usually used to estimate the niche of a species. Therefore, this study will also test whether the two databases provide the same environmental niche estimations. To sustain comparability to the results from Alhajeri & Fourcade (2019), the same taxonomic group, rodents, will be used. To counteract the uneven sampling in the GBIF dataset, the data will be thinned before proceeding with calculations, which is expected to increase the similarity between the two databases. The thinned GBIF data will be compared to the unthinned data to evaluate the method. The results from the multivariate method may shed light on differences not apparent in the univariate method and could give important additions to the debate about how different geographic range data influence macroecological studies and SDMs.

## Materials and methods

### Data extraction

The bioclimatic variables used in this study were downloaded from WorldClim (R. Hijmans, Cameron, Parra, Jones, & Jarvis, 2005) with a spatial resolution of 5 km<sup>2</sup>. The 19 bioclimatic variables are derived from monthly temperature and rainfall values and represent annual trends, seasonality and extreme limiting environmental factors. The explanation of each variable is presented in Table 1.

**Table 1:** Explanations of the 19 bioclimatic variables obtained from the WorldClim database (R. Hijmans et al., 2005).

|              |                                                            |              |                                                      |
|--------------|------------------------------------------------------------|--------------|------------------------------------------------------|
| <b>BIO1</b>  | Annual Mean Temperature                                    | <b>BIO11</b> | Mean Temperature of Coldest Quarter                  |
| <b>BIO2</b>  | Mean Diurnal Range (Mean of monthly (max temp - min temp)) | <b>BIO12</b> | Annual Precipitation                                 |
| <b>BIO3</b>  | Isothermality (BIO2/BIO7) (×100)                           | <b>BIO13</b> | Precipitation of Wettest Month                       |
| <b>BIO4</b>  | Temperature Seasonality (standard deviation ×100)          | <b>BIO14</b> | Precipitation of Driest Month                        |
| <b>BIO5</b>  | Max Temperature of Warmest Month                           | <b>BIO15</b> | Precipitation Seasonality (Coefficient of Variation) |
| <b>BIO6</b>  | Min Temperature of Coldest Month                           | <b>BIO16</b> | Precipitation of Wettest Quarter                     |
| <b>BIO7</b>  | Temperature Annual Range (BIO5-BIO6)                       | <b>BIO17</b> | Precipitation of Driest Quarter                      |
| <b>BIO8</b>  | Mean Temperature of Wettest Quarter                        | <b>BIO18</b> | Precipitation of Warmest Quarter                     |
| <b>BIO9</b>  | Mean Temperature of Driest Quarter                         | <b>BIO19</b> | Precipitation of Coldest Quarter                     |
| <b>BIO10</b> | Mean Temperature of Warmest Quarter                        |              |                                                      |

IUCN range maps were downloaded in Esri shapefile format (The IUCN Red List of Threatened Species, 2020) and loaded into R (R Core Team, 2018) with only the Rodentia order kept for all red list categories. For each species, the geographic polygon was rasterized using the raster library (R. J. Hijmans, 2020) at a resolution of 5 km<sup>2</sup>.

The GBIF occurrence data for the order Rodentia, was downloaded in a tab-delimited CSV format from GBIF (GBIF.org, 2020), and loaded into R (R Core Team, 2018). Before using the data, it was cleaned by excluding all records without coordinates, with less precision than 5 km around the survey location, with invalid coordinates, with equal longitude and latitude coordinates and coordinates equal to zero. The remaining occurrences were used when proceeding with further calculations. In parallel, the data was thinned as a method to counteract the clustering of occurrences, by only keeping one occurrence per climatic pixel. This resulted in two GBIF data frames – thinned and unthinned data.

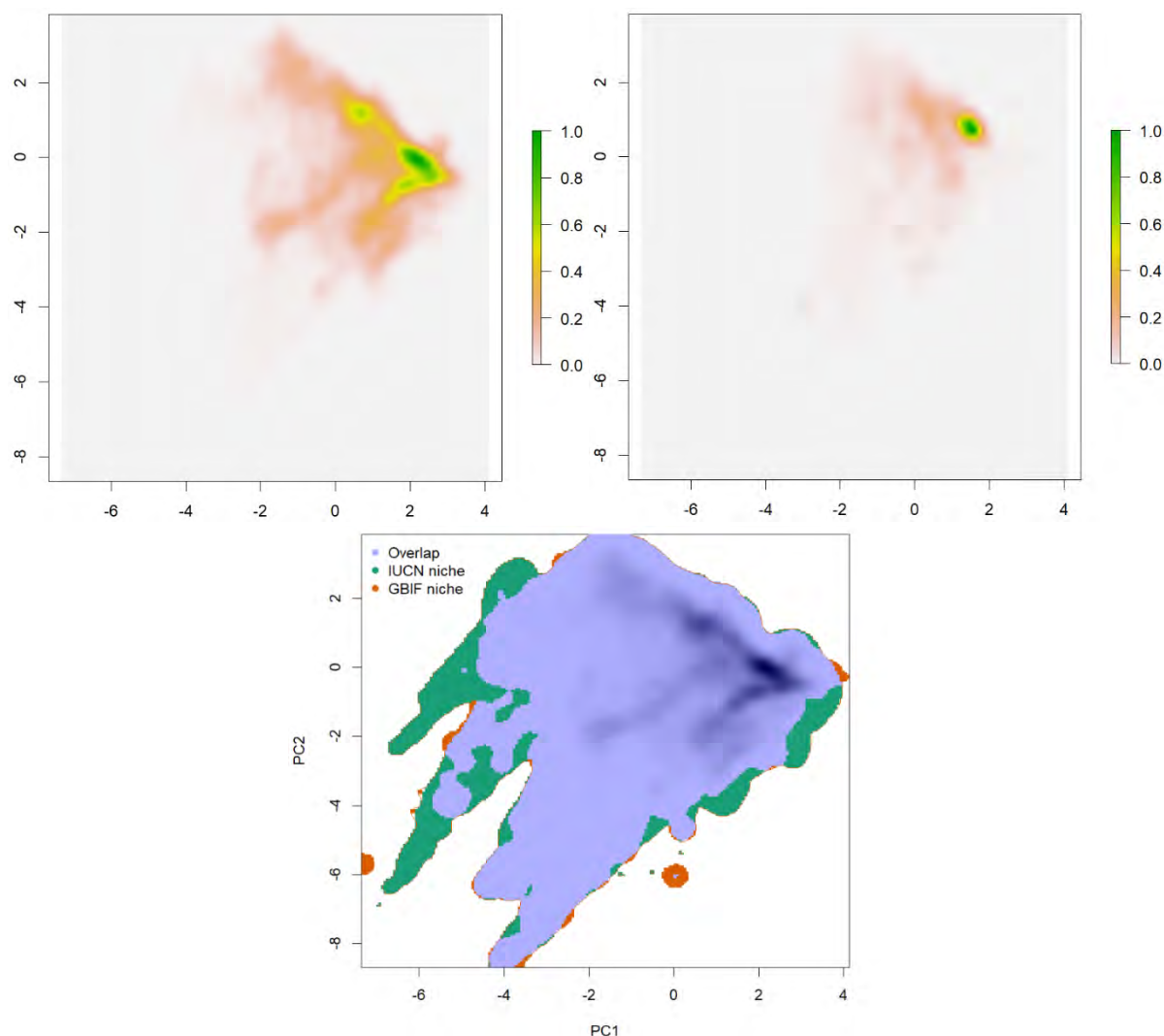
In both IUCN and GBIF data frames, only the species that were used in Alhajeri & Fourcade (2019) was kept to ensure comparability of the results. Among these, only species with at least 20 records for both data frames were considered. This resulted in 916 rodent species included in the IUCN and unthinned GBIF data frames while the thinned GBIF data frame contained 713 species. For every species, the values for each bioclimatic variable, where the species is present, were obtained from all datasets (IUCN, unthinned GBIF and thinned GBIF) and the mean, median and variance of the climatic variables were extracted (as done in Alhajeri and Fourcase 2019).

### **Data analysis**

Kendall's  $\tau$  coefficient was used to evaluate the association between the bioclimatic variables acquired from the three data frames across species. Kendall's correlation test is a nonparametric test that measures the rank correlation between two measured quantities (Kendall, 1938). Spearman's correlation test (Spearman, 1904) is a more commonly used correlation test for nonparametric data because of its superiority when there are any ties in the data (Puth, Neuhäuser, & Ruxton, 2015). Although, in data without ties, Kendall have been argued preferable because of its ability to generate narrower confidence intervals (Puth et al., 2015).

A Kendall's  $\tau$  test was performed between the mean environmental data for each bioclimatic variable obtained from either IUCN data and both the unthinned and thinned GBIF data, separately. This generated two Kendall's  $\tau$  of the mean and the same was also done for the median environmental data for each climatic variable. To explore whether the variability of environmental conditions obtained from the three data frames were the same, Kendall's  $\tau$  was also computed for the variance of each bioclimatic variable.

For the multivariate approach, a principal component analysis (PCA) of the world's values for the 19 bioclimatic variables was performed. The first and second axis explained 51% and 26% of the variance. These axes were chosen to define a two-dimensional environmental space describing the conditions existing on earth. From this, it was possible to get an estimate of the species realized niche by mapping the environmental conditions associated to a given species/data source into this two-dimensional space (Figure 3). This was done for all species based on data from the three data frames using the *ecospat* package (Broennimann, Di Cola, & Guisan, 2020) in R (R Core Team, 2018). The *ecospat* package was also used to overlay the estimations from the different data frames of the same species (Figure 3). To measure how big the overlap was, the Schoener's D index was used (Schoener, 1968). Schoener's D is a measure that ranges between 0 (no overlap) and 1 (identical estimations).



**Figure 3:** The estimated realized niche of *Peromyscus maniculatus* based on data from the International Union for Conservation of Nature (top-left) and the Global Biodiversity Information Facility (top-right), mapped in the two-dimensional environmental space of the bioclimatic values existing on earth. The overlap of the two estimations of the realized niche (bottom).

To evaluate if the method for reducing the spatial clustering bias in the GBIF dataset, spatial thinning, influenced the estimated niche of the species, Schoener's D from the overlap between IUCN and the two GBIF data frames were compared using a Wilcoxon test.

## Results

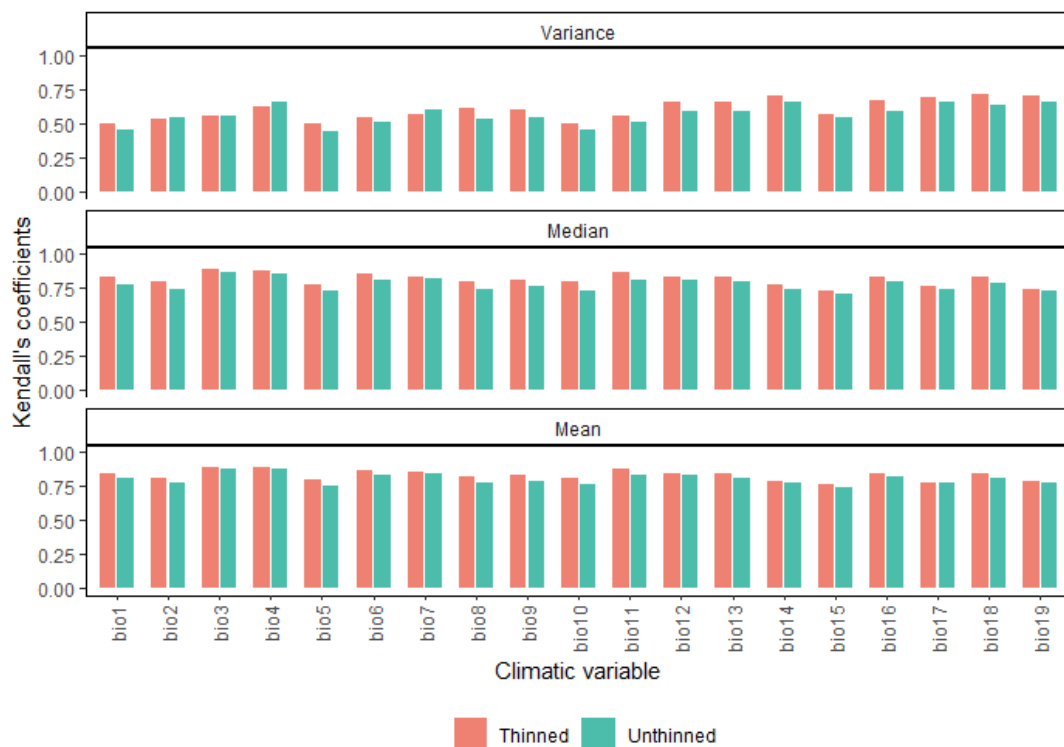
### Kendall's correlation between bioclimatic variables

There was high positive correlation between the mean bioclimatic variables obtained from the IUCN range maps and the GBIF occurrence data (Figure 4). Kendall's  $\tau$  ranged between 0.741-0.877 both when the unthinned and thinned GBIF data was used to obtain the bioclimatic variables.

The median bioclimatic variables from the two data sources also indicated high positive correlation, both when the thinned and the unthinned GBIF data was used (Figure 4).

Kendall's  $\tau$  ranged between 0.700-0.865 when using the unthinned data and between 0.728-0.888 using the thinned data.

The Kendall's  $\tau$  between the variance of the bioclimatic variables indicated a slightly lower positive correlation (Figure 4). The unthinned variance ranged between 0.442-0.663, while the thinned variance ranged between 0.498-0.715.

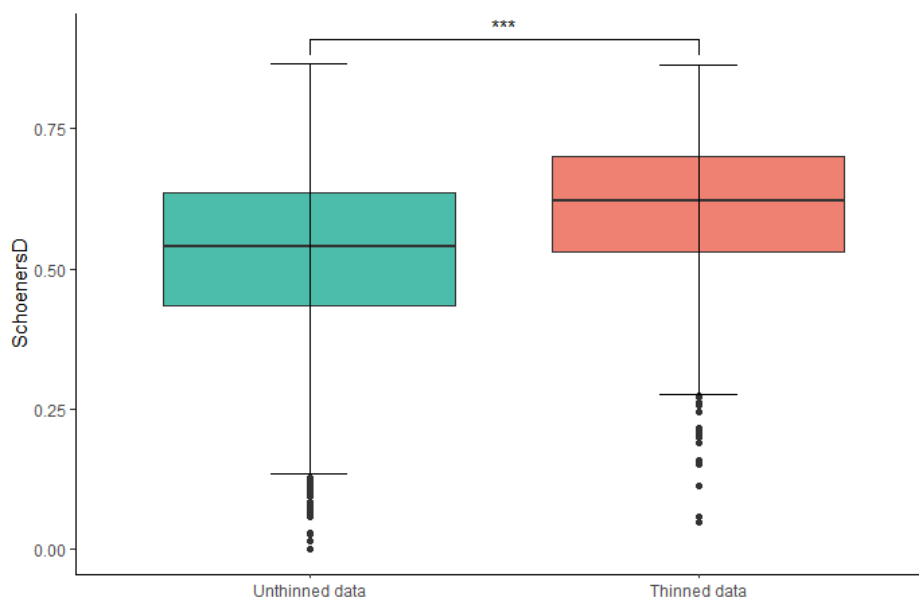


**Figure 4:** Three bar plots showing Kendall's  $\tau$  coefficient between the variance, median and mean of the bioclimatic variables obtained from IUCN range maps and GBIF occurrence data. Pink visualizes correlations using the thinned GBIF occurrence data while the blue visualizes the unthinned correlations.

### Niche overlap between the niche estimations from the different data sources

The results indicate that there is on average a 60% overlap between the two data sources and that some niches even showed close to zero overlap (Figure 5).

The results indicate a significant difference between the overlap obtained using the thinned and the unthinned GBIF data with a larger overlap with the former ( $p < 2.2e-16$ ,  $W = 220970$ ; Figure 5).



**Figure 5:** Boxplot of Schoener's D obtained by comparing the climatic niche of each species, obtained through IUCN range data and GBIF data. Blue visualizes overlap using unthinned GBIF data and Pink using thinned GBIF data. There is a significant difference between the Schoener's D obtained through the two methods.

## Discussion

Estimated species' ranges are used in macroecological studies, especially when training species distribution models (SDMs) (Warren et al., 2018) and when analysing macroecological patterns of different species (Fernanda Bonetti & Wiens, 2014; Savage & Vellend, 2015). These types of analyses are of great importance for ecological research and especially for conservation research and planning (Gaston & Fuller, 2009; Whittaker et al., 2005). The existing species distribution databases are undeniably valuable because of its high taxonomic, temporal, and spatial coverage and because of its accessibility. Therefore, knowledge about how biases affect the different data sources is essential to treat the existing data in a critical and impartial way. This study used two different approaches to test whether niche information derived from IUCN and GBIF provide similar estimations. The results from the estimations of the species' realized niche showed that the niches did not overlap more than 60% on average. This is a strong indication that these data bases do not give similar niche estimations and they should therefore be used with caution in this context. The results contradict the conclusions made by Alhajeri & Fourcade (2019) and show that when focusing on niche estimates within a two-dimensional space, the species-level environmental data diverge substantially. An example can be seen in Figure 3, where the estimated realized niche of the North American Deermouse (*P. maniculatus*) derived from the IUCN range map is larger than the realized niche derived from GBIF occurrences. This is a general trend in the data and the reason for the low overlap. IUCN range maps are known to overestimate the bioclimatic conditions experienced by the species (Hurlbert & Jetz, 2007) whereas the low spatial coverage of the GBIF data implies that this data tend to underestimate the climatic conditions experienced by the species.

The Kendall's  $\tau$  test performed between the mean and median of the bioclimatic variables, obtained through the different data sources, indicate that there is a high positive correlation.

Kendall's  $\tau$  was on average 0.79 for the median and 0.82 for the mean, across all the bioclimatic variables. These results are equivalent to the ones obtained by Alhajeri & Fourcade (2019) and does therefore suggest a similar conclusion, that the two databases provide comparable environmental information. For average values this is true but for some species this conclusion might not be appropriate. For example, the mean of the annual mean temperature (BIO1) within the estimated distribution range of the groundhog (*Marmota monax*) is estimated to be 2 °C with IUCN range maps and 9 °C using the GBIF occurrences. The reason why median and mean values of climatic variables are interesting is because they are often used as a method to get an estimate of the realised niche of a species (Li et al., 2009), to define the species' niche position (Castro-Insua, Gómez-Rodríguez, Wiens, & Baselga, 2018) and to calculate the distance between the mean habitat conditions used by the species and the mean habitat conditions in a specific study area (Dolédéc, Chessel, & Gimaret-Carpentier, 2000; Heino & Grönroos, 2014). It is also useful to simply condense the niche description to average values for the climatic conditions when studying niche patterns of species within a group or a clade of species (Lv, Xia, Ge, Wu, & Yang, 2016; Olalla-Tárraga et al., 2011) or when looking at the evolution of species' niches over long time scales (Jara-Arancio et al., 2014; Wiens, Kozak, & Silva, 2013).

To estimate a species niche breadth, variance of the climatic variables within the niche of the species can be used (Fernanda Bonetti & Wiens, 2014; Quintero & Wiens, 2013). Niche breadth is an important estimate when looking at species' ability to adapt to changes in biotic and abiotic factors such as changing climates, introduction of invasive species and land use changes. This study reveals that variance of the climatic data derived from IUCN range maps and GBIF occurrence data does not show as high correlation (Kendall's  $\tau$  0.6 on average). Because of the IUCN polygons being drawn in a simplified manner, they tend to include places with unsuitable habitats for the species, like mountain tops or in the middle of cities. This is in line with the perception that IUCN overestimates the climatic conditions experienced by species (Hurlbert & Jetz, 2007) which could be a reason for higher variance of some of the bioclimatic variables derived from IUCN. It is important to know how the two databases differ and how the biases affect these estimations. The biases are therefore very important to take in consideration when deciding on what resolution the data should be analysed in, both the climatic and the distribution data. This will directly influence variance of the estimated bioclimatic variables (Pearson & Dawson, 2003).

According to the results in this study, thinning the GBIF data made the estimated niches more like the niches derived from IUCN. These results are not surprising since the IUCN polygons already does not take the internal composition of species into account. As one of the well-known biases of the IUCN database, this can lead to estimations that are less precise because they include areas with directly unsuitable habitats for the species. Thinning the GBIF data can therefore be considered to decrease the sampling bias in the GBIF data frame but keep the precision in the spatial pattern. If the GBIF data is not thinned, some climatic conditions will receive overwhelming influence over the mean, median and variance but also for the niche estimations. Thinning the data is therefore a way to avoid some climatic conditions to be overrepresented because of sampling biases in the field. Thinned GBIF data could be a good data source to use in parallel with the IUCN polygons to estimate species' distributions.

Even though estimated niche positions using the two data sources might be similar (high correlations between the two databases regarding the median and the mean for all bioclimatic variables), the results from this study reveals that other niche estimations will differ considerably. Although, this study was only performed on rodent species, similar discrepancies could also exit for other taxa (Troia & McManamay, 2016). As an example, for birds, which tend to have a lot of volunteer recorders in the field (La Sorte & Somveille, 2020), the spatial coverage of the survey effort can be very uneven in GBIF (Huang, Lin, & Ji, 2020) which might cause even larger differences of niche estimations. The differences can also be strongly influenced by the spatial resolution of the data. The resolution of the data used in this study was 5 km<sup>2</sup> which is relatively small, especially when you consider how the IUCN range maps are drawn. Previous studies have advocated for using a coarser resolution because of the mismatch between the high resolution of climatic data and the lower resolution of distribution data (Hurlbert & Jetz, 2007; McPherson, Jetz, & Rogers, 2006). The resolution of the data should be decided depending on which species are studied, since home ranges of individuals can vary greatly between species. Studying available range maps for 834 different bird species, Hurlbert & Jetz (2007) showed that resolutions less than 200 km<sup>2</sup> overestimated the occupancy area of individual species and mischaracterized the spatial pattern of the species richness. In future studies it would therefore be interesting to study the relationship of the environmental estimates of rodents at a coarser resolution to see how the estimations differ. It would also be interesting to investigate how these relationships appear for other taxa than rodents.

The results from this study reveals that for some applications (univariate assessments) the two databases will not differ much, but for other applications (niche breadth and multivariate assessments) it will differ considerably. Environmental niche estimations from IUCN and GBIF are frequently used to produce SDMs to forecast future changes in species distributions (Thuiller, 2004) or species extinction risks (Rowland, Davison, & Graumlich, 2011). Since SDMs are based on the niche concept and since niche estimates evidently can vary between the two databases, distribution change forecasts might vary strongly depending on which database is used. This would result in SDMs showing divergent conclusions when trained with data from the different sources (Björklund, Lindelöw, & Schroeder, 2016; Fourcade, 2016; Fourcade, Engler, Besnard, Rödder, & Secondi, 2013). A general conclusion from this study is that one database cannot be chosen over the other since they both suffer many biases, and it is not clear which database is more reliable. Therefore, it would be safer to always consider both databases and interpret the results in regard of their biases. If both databases provide similar results, it will strengthen the conclusions, while divergent results might be a way to detect a reality somewhere in between.

With the increasing loss of biodiversity during the last century (Butchart et al., 2010; Dirzo et al., 2014), it is important to improve the knowledge about the environmental niches used by species, both today and in the future. Appropriate design of SDMs is therefore important for Red list assessments (Breiner, Guisan, Nobis, & Bergamini, 2017) and for prioritizing conservation action (Whittaker et al., 2005).

## Acknowledgements

First, I would like to thank my supervisor Mathieu Chevalier for his guidance and support, throughout the whole project and with his feedback on this text. He also helped me by generating the explanatory figures 1-3. I would also like to thank Antoine Guisan and the whole research group for the feedback and discussion addressed on the topic of my project. This project is based on data retrieved from GBIF, IUCN and WorldClim, which would not be accessible without the many contributions from both professionals and volunteers.

## References

- Alhajeri, B., & Fourcade, Y. (2019). High correlation between species-level environmental data estimates extracted from IUCN expert range maps and from GBIF occurrence data. *Journal of Biogeography*, 46(7), 1329–1341. doi: 10.1111/jbi.13619
- Beck, J., Böller, M., Erhardt, A., & Schwanghart, W. (2014). Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecological Informatics*, 19, 10–15. doi: 10.1016/j.ecoinf.2013.11.002
- Björklund, N., Lindelöw, Å., & Schroeder, L. M. (2016). Erroneous conclusions about current geographical distribution and future expansion of forest insects in Northern Sweden: Comments on Hof and Svahlin (2015). *Scandinavian Journal of Forest Research*, 31(1), 126–127. doi: 10.1080/02827581.2015.1086018
- Boria, R. A., Olson, L. E., Goodman, S. M., & Anderson, R. P. (2014). Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecological Modelling*, 275, 73–77. doi: 10.1016/j.ecolmodel.2013.12.012
- Boria, R., & Blois, J. (2018). The effect of large sample sizes on ecological niche models: Analysis using a North American rodent, *Peromyscus maniculatus*. *Ecological Modelling*, 386, 83–88. doi: 10.1016/j.ecolmodel.2018.08.013
- Breiner, F., Guisan, A., Nobis, M., & Bergamini, A. (2017). Including environmental niche information to improve IUCN Red List assessments. *Diversity and Distributions*, 23(5), 484–495. doi: <https://doi.org/10.1111/ddi.12545>
- Broennimann, O., Di Cola, V., & Guisan, A. (2020). ecospat: Spatial Ecology Miscellaneous Methods. *R Package Version 3.1*. Retrieved from <https://CRAN.R-project.org/package=ecospat>
- Brown, J., Stevens, G., & Kaufman, D. (1996). The geographic range: Size, Shape, Boundaries, and Internal Structure. *Annual Review of Ecology and Systematics*, 27(1), 597–623. doi: 10.1146/annurev.ecolsys.27.1.597
- Butchart, S. H. M., Walpole, M., Collen, B., van Strien, A., Scharlemann, J., Almond, R., ... Watson, R. (2010). Global Biodiversity: Indicators of Recent Declines. *Science*, 328(5982), 1164–1168. doi: 10.1126/science.1187512

- Castro-Insua, A., Gómez-Rodríguez, C., Wiens, J., & Baselga, A. (2018). Climatic niche divergence drives patterns of diversification and richness among mammal families. *Scientific Reports*, 8(1), 8781. doi: 10.1038/s41598-018-27068-y
- Curtis, C., & Bradley, B. (2016). Plant Distribution Data Show Broader Climatic Limits than Expert-Based Climatic Tolerance Estimates. *Plos One*, 11(11), e0166407. doi: 10.1371/journal.pone.0166407
- 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(2), 73–77. doi: 10.1023/A:1009690919835
- Dickinson, J., Zuckerberg, B., & Bonter, D. (2010). Citizen Science as an Ecological Research Tool: Challenges and Benefits. *Annual Review of Ecology, Evolution, and Systematics*, 41(1), 149–172. doi: 10.1146/annurev-ecolsys-102209-144636
- Dirzo, R., Young, H. S., Galetti, M., Ceballos, G., Isaac, N. J. B., & Collen, B. (2014). Defaunation in the Anthropocene. *Science*, 345(6195), 401–406. doi: 10.1126/science.1251817
- Dolédec, S., Chessel, D., & Gimaret-Carpentier, C. (2000). Niche Separation in Community Analysis: A New Method. *Ecology*, 81(10), 2914–2927. doi: [https://doi.org/10.1890/0012-9658\(2000\)081\[2914:NSICAA\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2000)081[2914:NSICAA]2.0.CO;2)
- Feng, X., Castro, M., McBee, K., & Papes, M. (2017). Hiding in a Cool Climatic Niche in the Tropics? An Assessment of the Ecological Biogeography of Hairy Long-Nosed Armadillos (*Dasypus pilosus*). *Tropical Conservation Science*, 10. doi: 10.1177/1940082917697249
- Fernanda Bonetti, M., & Wiens, J. (2014). Evolution of climatic niche specialization: A phylogenetic analysis in amphibians. *Proceedings of the Royal Society B: Biological Sciences*, 281(1795), 20133229. doi: 10.1098/rspb.2013.3229
- 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(11), 1303–1313. doi: 10.1111/geb.12201
- Fourcade, Y. (2016). Comparing species distributions modelled from occurrence data and from expert-based range maps. Implication for predicting range shifts with climate change. *Ecological Informatics*, 36, 8–14. doi: 10.1016/j.ecoinf.2016.09.002
- Fourcade, Y., Engler, J. O., Besnard, A. G., Rödder, D., & Secondi, J. (2013). Confronting expert-based and modelled distributions for species with uncertain conservation status: A case study from the corncrake (*Crex crex*). *Biological Conservation*, 167, 161–171. doi: 10.1016/j.biocon.2013.08.009
- Fourcade, Y., Engler, J. O., Rödder, D., & Secondi, J. (2014). Mapping Species Distributions with MAXENT Using a Geographically Biased Sample of Presence Data: A

- Performance Assessment of Methods for Correcting Sampling Bias. *PLoS ONE*, 9(5). doi: 10.1371/journal.pone.0097122
- Gaikwad, J., Wilson, P., & Ranganathan, S. (2011). Ecological niche modeling of customary medicinal plant species used by Australian Aborigines to identify species-rich and culturally valuable areas for conservation. *Ecological Modelling*, 222(18), 3437–3443. doi: 10.1016/j.ecolmodel.2011.07.005
- Gaston, K., & Fuller, R. (2009). The sizes of species' geographic ranges. *Journal of Applied Ecology*, 46(1), 1–9. doi: <https://doi.org/10.1111/j.1365-2664.2008.01596.x>
- GBIF.org. (2020). GBIF Occurrence Download (<https://doi.org/10.15468/dl.z4v3cq>)
- Guisan, A., & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. *Ecology Letters*, 8(9), 993–1009. doi: <https://doi.org/10.1111/j.1461-0248.2005.00792.x>
- Heino, J., & Grönroos, M. (2014). Untangling the relationships among regional occupancy, species traits, and niche characteristics in stream invertebrates. *Ecology and Evolution*, 4(10), 1931–1942. doi: <https://doi.org/10.1002/ece3.1076>
- Herk, K. M., Skidmore, A., & Fahr, J. (2017). Macroecological conclusions based on IUCN expert maps: A call for caution. *Global Ecology and Biogeography*, 26(8), 930–941. doi: 10.1111/geb.12601
- Hijmans, R., Cameron, S., Parra, J., Jones, P., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25(15), 1965–1978. doi: <https://doi.org/10.1002/joc.1276>
- Hijmans, R. J. (2020). raster: Geographic Data Analysis and Modeling. *R Package Version 3.1-5*. Retrieved from <https://CRAN.R-project.org/package=raster>
- Huang, X., Lin, C., & Ji, L. (2020). The persistent multi-dimensional biases of biodiversity digital accessible knowledge of birds in China. *Biodiversity and Conservation*, 29(11–12), 3287–3311. doi: 10.1007/s10531-020-02024-3
- Hurlbert, A., & Jetz, W. (2007). Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. *Proceedings of the National Academy of Sciences*, 104(33), 13384–13389. doi: 10.1073/pnas.0704469104
- Isaac, N., van Strien, A., August, T., de Zeeuw, M., & Roy, D. (2014). Statistics for citizen science: Extracting signals of change from noisy ecological data. *Methods in Ecology and Evolution*, 5(10), 1052–1060. doi: <https://doi.org/10.1111/2041-210X.12254>
- IUCN Red List. (2020). The IUCN Red List of Threatened Species. Retrieved April 7, 2020, from IUCN Red List of Threatened Species website: <https://www.iucnredlist.org/en>
- Jara-Arancio, P., Arroyo, M., Guerrero, P., Hinojosa, L., Arancio, G., & Méndez, M. (2014). Phylogenetic perspectives on biome shifts in *Leucocoryne* (Alliaceae) in relation to

- climatic niche evolution in western South America. *Journal of Biogeography*, 41(2), 328–338. doi: <https://doi.org/10.1111/jbi.12186>
- Kendall, M. G. (1938). A New Measure of Rank Correlation. *Biometrika*, 30(1/2), 81–93. doi: 10.2307/2332226
- La Sorte, F., & Somveille, M. (2020). Survey completeness of a global citizen-science database of bird occurrence. *Ecography*, 43(1), 34–43. doi: 10.1111/ecog.04632
- Li, J., He, Q., Hua, X., Zhou, J., Xu, H., Chen, J., & Fu, C. (2009). Climate and history explain the species richness peak at mid-elevation for Schizothorax fishes (Cypriniformes: Cyprinidae) distributed in the Tibetan Plateau and its adjacent regions. *Global Ecology and Biogeography*, 18(2), 264–272. doi: <https://doi.org/10.1111/j.1466-8238.2008.00430.x>
- Lv, X., Xia, L., Ge, D., Wu, Y., & Yang, Q. (2016). Climatic niche conservatism and ecological opportunity in the explosive radiation of arvicoline rodents (Arvicolinae, Cricetidae). *Evolution*, 70. doi: 10.1111/evo.12919
- Martínez-Méndez, N., Aguirre-Planter, E., Eguiarte, L., & Jaramillo-Correa, J. P. (2016). Modelado de nicho ecológico de las especies del género *Abies* (Pinaceae) en México: Algunas implicaciones taxonómicas y para la conservación. *Botanical Sciences*, 94(1), 5–24. doi: 10.17129/botsci.508
- McPherson, J., Jetz, W., & Rogers, D. (2006). Using coarse-grained occurrence data to predict species distributions at finer spatial resolutions—Possibilities and limitations. *Ecological Modelling*, 192(3–4), 499–522. doi: 10.1016/j.ecolmodel.2005.08.007
- Olalla-Tárraga, M., McInnes, L., Bini, L., Diniz-Filho, J., Fritz, S., Hawkins, B., ... Purvis, A. (2011). Climatic niche conservatism and the evolutionary dynamics in species range boundaries: Global congruence across mammals and amphibians. *Journal of Biogeography*, 38(12), 2237–2247. doi: 10.1111/j.1365-2699.2011.02570.x
- Pearson, R., & Dawson, T. (2003). Predicting the impacts of climate change on the distribution of species: Are bioclimate envelope models useful? *Global Ecology and Biogeography*, 12(5), 361–371. doi: <https://doi.org/10.1046/j.1466-822X.2003.00042.x>
- Puth, M.-T., Neuhäuser, M., & Ruxton, G. (2015). Effective use of Spearman's and Kendall's correlation coefficients for association between two measured traits. *Animal Behaviour*, 102, 77–84. doi: 10.1016/j.anbehav.2015.01.010
- Quintero, I., & Wiens, J. (2013). What determines the climatic niche width of species? The role of spatial and temporal climatic variation in three vertebrate clades. *Global Ecology and Biogeography*, 22(4), 422–432. doi: <https://doi.org/10.1111/geb.12001>
- R Core Team. (2018). R: A language and environment for statistical computing (Version 3.5.2) [R, X86\_64, mingw32]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>

- Romo, H., García-Barros, E., & Lobo, J. (2006). Identifying recorder-induced geographic bias in an Iberian butterfly database. *Ecography*, 29(6), 873–885. doi: <https://doi.org/10.1111/j.2006.0906-7590.04680.x>
- Rowland, E., Davison, J., & Graumlich, L. (2011). Approaches to Evaluating Climate Change Impacts on Species: A Guide to Initiating the Adaptation Planning Process. *Environmental Management*, 47(3), 322–337. doi: 10.1007/s00267-010-9608-x
- Sagarin, R., & Gaines, S. (2002). The ‘abundant centre’ distribution: To what extent is it a biogeographical rule? *Ecology Letters*, 5(1), 137–147. doi: <https://doi.org/10.1046/j.1461-0248.2002.00297.x>
- Savage, J., & Vellend, M. (2015). Elevational shifts, biotic homogenization and time lags in vegetation change during 40 years of climate warming. *Ecography*, 38(6), 546–555. doi: <https://doi.org/10.1111/ecog.01131>
- Schipper, J., Chanson, J., Chiozza, F., Cox, N., Hoffmann, M., Katariya, V., ... Young, B. E. (2008). The status of the world’s land and marine mammals: Diversity, threat, and knowledge. *Science*, 322(5899), 225–230. doi: 10.1126/science.1165115
- Schmeller, D., Henry, P.-Y., Julliard, R., Gruber, B., Clobert, J., Dziack, F., ... Henle, K. (2009). Advantages of volunteer-based biodiversity monitoring in Europe. *Conservation Biology: The Journal of the Society for Conservation Biology*, 23(2), 307–316. doi: 10.1111/j.1523-1739.2008.01125.x
- Schoener, T. (1968). The Anolis Lizards of Bimini: Resource Partitioning in a Complex Fauna. *Ecology*, 49(4), 704–726. doi: 10.2307/1935534
- Snäll, T., Kindvall, O., Nilsson, J., & Pärt, T. (2011). Evaluating citizen-based presence data for bird monitoring. *Biological Conservation*, 144(2), 804–810. doi: 10.1016/j.biocon.2010.11.010
- Soberón, J. (2007). Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters*, 10(12), 1115–1123. doi: 10.1111/j.1461-0248.2007.01107.x
- Spearman, C. (1904). The Proof and Measurement of Association between Two Things. *The American Journal of Psychology*, 15(1), 72–101. doi: 10.2307/1412159
- The IUCN Red List of Threatened Species. (2020). Spatial Data Download. Retrieved December 7, 2020, from Spatial data and mapping resources website: <https://www.iucnredlist.org/en>
- Thuiller, W. (2004). Patterns and uncertainties of species’ range shifts under climate change. *Global Change Biology*, 10(12), 2020–2027. doi: <https://doi.org/10.1111/j.1365-2486.2004.00859.x>
- Tiago, P., Pereira, H., & Capinha, C. (2017). Using citizen science data to estimate climatic niches and species distributions. *Basic and Applied Ecology*, 20, 75–85. doi: 10.1016/j.baae.2017.04.001

- Troia, M., & McManamay, R. (2016). Filling in the GAPS: Evaluating completeness and coverage of open-access biodiversity databases in the United States. *Ecology and Evolution*, 6(14), 4654–4669. doi: <https://doi.org/10.1002/ece3.2225>
- Vetaas, O. (2002). Realized and potential climate niches: A comparison of four *Rhododendron* tree species. *Journal of Biogeography*, 29(4), 545–554. doi: <https://doi.org/10.1046/j.1365-2699.2002.00694.x>
- Warren, R., Price, J., Graham, E., Forstenhaeusler, N., & VanDerWal, J. (2018). The projected effect on insects, vertebrates, and plants of limiting global warming to 1.5°C rather than 2°C. *Science*, 360(6390), 791–795. doi: 10.1126/science.aar3646
- Whittaker, R., Araújo, M., Jepson, P., Ladle, R., Watson, J., & Willis, K. (2005). Conservation Biogeography: Assessment and prospect. *Diversity and Distributions*, 11(1), 3–23. doi: <https://doi.org/10.1111/j.1366-9516.2005.00143.x>
- Wiens, J., Kozak, K., & Silva, N. (2013). Diversity and Niche Evolution Along Aridity Gradients in North American Lizards (phrynosomatidae). *Evolution*, 67(6), 1715–1728. doi: <https://doi.org/10.1111/evo.12053>