Journal of Urban Ecology, 2022, 1–10

https://doi.org/10.1093/jue/juac002
Research Article

Data from public and governmental databases show that a large proportion of the regional animal species pool occur in cities in Germany

Fabio S.T. Sweet [1,*], Beate Apfelbeck,[1,2] Maximilian Hanusch [1,2]
Cynthia Garland Monteagudo[1] and Wolfgang W. Weisser[1]

[1]Terrestrial Ecology Research Group, Department of Life Science Systems, School of Life Sciences, Technical University of Munich, Hans-Carl-von-Carlowitz-Platz 2, Freising 85354, Germany and [2]Department of Environment and Biodiversity, Paris-Lodron-Universität Salzburg, Hellbrunnerstraße 34, Salzburg 5020, Austria

*Corresponding author. E-mail: fabio.sweet@tum.de

Submitted: 1 July 2021; **Received (in revised form):** 4 January 2022. **Accepted:** 13 January 2022

Abstract

Cities have been shown to be biodiverse, but it is unclear what fraction of a regional species pool can live within city borders and how this differs between taxa. Among animals, most research has focused on a few well-studied taxa, such as birds or butterflies. For other species, progress is limited by the paucity of data. We used species occurrence data for 11 taxa and 23 German cities from the Global Biodiversity Information Facility (GBIF) and the different German states, in a 50-km buffer around the city centre, to investigate what proportion of species of the regional species pools also occur in cities. While data could be obtained for all cities from GBIF, state databases only provided data for a subset of cities. Sample coverage of data from GBIF was higher across all taxa than of the state databases. For each database and taxon, we analysed (i) all cities where the number of occurrences of a taxon was >50 and (ii) only those cities where additionally sample coverage was >0.85. Across all taxa studied on average, 44.9 ± 7.2% (GBIF) and 40.8 ± 9.6% (German states) of the species of the regional species pool were also found in cities. When all cities were considered together, more than 76% of all species occurred within city borders. Our results show that German cities harbour a large part of the regional diversity of different taxa when city borders rather than the city centre is considered. This opens up ample opportunities for conservation and for fostering human–nature relationships.

Key words: biodiversity, biodiversity databases, cities, GBIF, Germany, species richness, mammals, birds, reptiles, amphibians, coleoptera, diptera, hemiptera, hymenoptera, lepidoptera, orthoptera, spiders

Introduction

Urban areas have been predicted to rapidly expand within the next decades (Chen et al. 2020). By replacing natural and agricultural areas with built environments, urbanization profoundly changes land cover and use. Concomitantly, many aspects of the abiotic environment are changed, such as climate and hydrology of an area, with consequences for the biodiversity of the region (Grimm et al. 2008). It has been postulated that cities act as environmental filters that only allow a percentage of the species of the regional species pool to live within them. Certain functional traits may be favoured in urban environments, which leads to shifts in species composition between rural and urban environments (Aronson et al. 2016). Because the same subset of traits is favoured across the globe and because not all species of the regional species pool can live in cities, urbanization is seen to promote global biodiversity loss and global and regional homogenization of animals and plants (McKinney 2006).

Birds have been studied particularly well and studies suggest that species density and phylogenetic diversity is reduced in urban areas (Aronson et al. 2014; Morelli et al. 2016; Ibáñez-Álamo et al. 2017). Furthermore, urban dwellers are characterized by similar life-history traits and are often species with a wide geographical range (Bonier et al. 2007; Guetté et al. 2017). Similarly, in arthropods, urbanization leads to decreased species richness, higher species evenness and community shifts towards species with certain traits (Knop 2016; Piano et al. 2020). For example, higher temperatures in urban areas seem to favour smaller species within different arthropod taxa (Merckx et al. 2018). However, studies on arthropods usually only cover few taxa within these species-rich groups (Fenoglio et al. 2020). The effects of urbanization on other taxonomic groups have been studied even less well, but in general patterns appear to be similar, i.e. urban areas support fewer species and these have similar traits (e.g. bats: Duchamp and Swihart 2008; Jung and Threlfall 2018).

While these studies confirm the negative effects of urbanization on biodiversity globally, studies in Europe, where urban areas have a long history, also show that a large proportion of bird species found in a region may also occur in cities (Ferenc et al. 2014; Guetté et al. 2017). Despite selection for similar traits in city species, only a few global species exist and a variable part of urban biota continues to reflect the regional species pool (e.g. birds and plants: Aronson et al. 2014). A recent meta-analysis of urban gradient studies on terrestrial animals did not confirm that species richness of different animal taxa is consistently lower in urban areas (Saari et al. 2016). In addition, studies that relate urban biodiversity to urban features suggest that negative effects of urban areas on biodiversity depend on the structure of a city, i.e. biodiversity within cities depends on the availability of appropriate habitat features, such as green spaces and vegetation (Beninde et al. 2015; Threlfall et al. 2016). In consequence, cities may harbour a high species diversity. In fact, cities commonly are richer in species than the increasingly intensified agricultural areas surrounding them (Turrini and Knop 2015).

A general challenge for studies on the occurrence of species in cities is the availability of data. Individual transect or plot-based studies provide high-quality data, yet there is a limit as to how many plots a single study can assess. Today, with increased digitalization and the advent of large, publicly available databases, more data on species occurrences have become available. Large-scale monitoring efforts with many participants have the advantage that there are generally more data points collected over a wider spatial area. Birds have been most extensively studied through large-scale monitoring efforts (Greenwood 2007), because of the many volunteers taking part in standard counts such as campaigns from the British Trust for Ornithology (Harris et al. 2020), or the German bird monitoring scheme (Schwarz and Flade 2000). Besides the traditional large-scale monitoring schemes, there is now an increasing amount of data coming from newer citizen science projects such as eBird (Sullivan et al. 2009) and iNaturalist (www.inaturalist.org). In addition, there has been a worldwide effort to standardize data storage and make science-grade species occurrence data readily available through the Global Biodiversity Information Facility (GBIF: The Global Biodiversity Information Facility 2020). Many organizations and institutes, including NGOs such as the German NABU, iNaturalist, the Cornell lab of Ornithology, and many museums, are already supplying standardized data to this database. In addition to these public databases, there are also governmental databases used for conservation planning. In Germany, the 16 German states have the responsibility to set up and maintain such databases. Data entering these databases are a mixture of data obtained from government-financed monitoring efforts, often restricted to a particular taxon or a particular location, data obtained in the framework of environmental impact assessments, but also data that citizens or NGOs contribute to these databases. Currently, there is no clear relationship between GBIF data and the dataset of the German states, as any agreement on data sharing depends on the particular German state and as a rule, government data are not yet supplied to GBIF. Both datasets are commonly used, albeit for different purposes. The German states' data are applied for nature conservation within Germany, while GBIF data are commonly used within science. Against this backdrop, it is important to consider whether these two datasets confer similar results or not, as this could have implications for practice and policy.

Here, we take advantage of the databases of GBIF and the German states to test the idea that, depending on the taxonomic group, cities may contain a high percentage of the regional species pool. We compared the species richness of 23 cities in Germany with the species richness of the surrounding area, within a 50-km radius around the city centre. In our study, the regional species pool includes all species within the 50-km radius but outside the city borders. Following the conceptual framework of Aronson et al. (2016), we assume that the majority of species within a city recruits from the region surrounding a city, but that not all species are able to colonize it. Further, some species may have been actively introduced by humans (in particular in the case of plants), and thus may not be represented in the regional species pool. The regional species pool is determined by biogeography, but also past and present land use and therefore, encompasses both natural but also cultivated areas and settlements. In our study, we include data of 11 animal taxa. Our comparative analysis of the two databases GBIF and the German States also provides recommendations for results based on the current state of publicly available data.

Methods

Raw data collection

We selected 23 cities in Germany with at least one city from each of the 16 states of Germany (Supplementary Appendix A, Fig. A1). Fifteen of these cities were chosen because they contained a project site of another research project on promoting animals in cities (Weisser and Hauck 2017. Eight further cities with >100 000 inhabitants were selected to have at least one city represented from each German state. Using R (R Development Core Team 2008) within RStudio IDE (Rstudio Team 2015), spatial data of the political boundaries of these were obtained from the Database of Global Administrative Areas (University of California Berkely 2018) and merged into one shapefile. Separate shapefiles were made for the administrative borders of the cities, and for a 50-km buffer around the centre of each city.

Data were requested and gathered from two main sources, namely GBIF and the individual German states. We included data from 11 taxa, including vertebrates (Amphibia, Aves, Mammalia, Reptilia), insects (Coleoptera, Diptera, Hemiptera, Hymenoptera, Lepidoptera, Orthoptera) and Arachnids (Araneae).

Animal occurrence data from 1980 to 2018 were downloaded separately for each taxon from the GBIF website: birds on 19 August 2020; mammals on 31 August 2020; amphibians, reptiles and insects on 6 September 2020; and arachnids on 7 September 2020 (GBIF.org 2020a,b,c,d,e,f). The year 1980 was chosen as a starting point because the number of animal observation records in Germany strongly increased around that period of time (data not shown). Of each taxonomic group, the occurrence data for all of Germany were downloaded. These were then intersected with 50-km radiuses around the city centres of the 23 selected German cities.

Data from the German states were requested in writing for the period 1935–2018. Only data from 1980 onwards were used in the analyses. Thirteen states supplied observation data within 50 km of predetermined locations within the city (BUE 2018; LANUV NRW 2018; LAU 2018; LfU Bayern 2018; LfU Rheinland-Pfalz 2018; LfULG 2018; LLUR 2018; LUBW 2018; LUNG 2018; LVGL 2018; NKWKN 2018; Stiftung Naturschutz Berlin 2018; TLUG 2018; VSWFFM 2018). Berlin and Hamburg supplied observation data within their city-state border. Thüringen supplied observation data within a 20-km radius, and Brandenburg, Bremen and Baden-Württemberg did not provide observation data, for several different reasons including the absence of a functional database, low number of species records, or work overload. In cases where the 50-km buffer of a city extended into a different state, observation data from the region around a city that was not in its respective state might be unavailable and we worked with all data that were available to us (Supplementary Appendix A, Figs. A2 and A3). Overall, 18 of the cities could be investigated using data from the German states.

Data processing

All analyses were conducted in R (R Development Core Team 2008) using RStudio IDE (Rstudio Team 2015). To download data from OpenStreetMap (OpenStreetMap Contributors 2021), we used QuickOSM (Trimaille 2020) in QGIS (QGIS Development Team 2021) to create shapefiles of spatial data from OpenStreetMap.

Raw data from species databases often had faulty coordinates or species names that were misspelled or outdated. Data also included a number of different types of species records (e.g. human observations, machine observations and fossil records), of which only a part represented field observations. We addressed these issues by applying the following steps to clean the data (see script in Supplementary Appendix H).

First, records that were erroneously georeferenced to the country as a whole (i.e. the species occurs 'somewhere in Germany'), that were close to biodiversity institutions with collections of live or dead animals such as museums or zoos, that were not within Germany based on the coordinates or that had other spatial issues were removed with the R-package 'CoordinateCleaner' (Zizka et al. 2019).

Second, we removed occurrence records other than field observations from the data, such as museum specimens, as 'CoordinateCleaner' does not have all institutions listed. To make sure that observations within zoos were removed, observations where the metadata indicated that they were in locations that included the words 'zoo', 'tiergarten' and 'botanisch' were removed. In addition, we created a shapefile using QuickOSM (Trimaille 2020) in QGIS (QGIS Development Team 2021) from the OpenStreetMap (OpenStreetMap Contributors 2021) and removed locations and polygons of all zoos and zoo-like areas in Germany.

Third, observations where the species was unknown or where the name indicated a hybrid species were removed from the data. Subspecies were simplified to their binomial species name, and for the vertebrate groups (amphibians, birds, mammals and reptiles), synonym names were changed to the currently accepted scientific names manually. To further corroborate the list of species analysed, we used the German red lists. In Germany, a red list is always made for an entire taxon at a time, and hence includes a reviewed list of all species of this taxon that occur in Germany. Because of the large number of species that need to be assessed, red lists are normally made for lower taxonomic levels (i.e. not for 'insects'). We used the red lists for birds (Grüneberg et al. 2015), mammals (Meinig et al. 2020), amphibians (Kühnel et al. 2009), reptiles (Rote-Liste-Gremium Amphibien und Reptilien 2020), Araneae (Blick et al. 2016), and Orthoptera (Maas et al. 2011). For these taxa, species that were present in the red list were kept for further analysis, independent of the frequency of occurrence. Of the species that we did not find in the red list, those with at least 50 observations in all of Germany were kept for further analysis. This selection was done on the basis of all of the observations in Germany (GBIF), or on the complete dataset (states' data). Because Diptera and Hemiptera had in total only 1 and 47 observations, respectively, across all cities and regions in the German state data, we excluded these taxa from the state dataset. To account for severe undersampling, if a city had fewer than 51 observations for a particular taxon, that city was excluded for further analysis.

After data processing, 5.568.438 datapoints in the GBIF dataset and 2.623.835 datapoints in the federal state dataset were used for further analysis (Supplementary Appendix B, Table B1). Note if two or more of the 50-km buffers around the cities overlapped, observations that fell in two or more buffers were used for the analysis of each city independently and were thus used multiple times.

We defined the regional species pool as those species that occur outside the focal city borders but within the 50-km buffer. Thus, the regional species pool did not include species that only occurred within the city borders.

Data analysis

Analyses were done for each of the taxa individually and for all of them combined. GBIF and German states datasets were analysed separately. Taxa and cities were not equally well sampled, and sampling effort differed between databases. We calculated both Chao's measure of sample coverage (Chao and Jost 2012), and rarefaction curves for each taxon, city/region and data source separately (Supplementary Appendix C, Fig. C1) with the iNEXT R-package (Hsieh et al. 2016). We used a sample coverage of 0.85 to consider a taxon to be well sampled for a particular city, slightly higher than found in other analyses (e.g. Haack et al. 2021; Leveau 2021). All further analyses were performed both before and after applying the sample coverage filter.

We compared the lists of species observed within the city borders to the lists of species observed in the surrounding areas. We calculated (i) the number of species occurring only in the region, (ii) the number of species occurring both in the region and in the city and (iii) the number of species only found within the city. This analysis was performed for every taxon, city and dataset separately. We investigated whether there was a significant difference between the mean species richness of cities and the surrounding regions, for each taxon and each dataset, with the use of paired t-tests. As cities differed in their species inventory, we also calculated the same measures for all cities combined and compared this to all regions combined. We did not correct for city size or the size of the surrounding regions. We also did not exclude areas of smaller towns and villages from the regional buffer, as there are very few unmodified areas in Germany, and any decision to exclude areas in the region would still result in the remaining area being under the influence of humans. Additionally, as there is the potential of movement of animals between urbanized areas, it provides a more complete city-to-region comparison with the inclusion of similar ecosystems from the surroundings.

We ran generalized additive models (GAM) using the *mgcv* R-package (Wood 2021) to analyse the relationship between the percentage of species of the 50-km buffer that was found in a city against the percentage of all observations that were made within the city. For details on model settings, please see Supplementary Appendix C. Plots were drawn using the ggplot2 (Wickham 2016) and Patchwork (Pedersen 2019) R-packages.

Results

Sample coverage and species richness

Of 253 possible taxon × city combinations within the GBIF dataset (23 cities × 11 taxa), 141 combinations fulfilled the criterion of >50 observations within the city. Of these, 120 combinations also had a sampling coverage of >0.85 (Supplementary Appendix C). Within the German states dataset, only 62 taxon × city combinations fulfilled the criterion of >50 observations within the city, and of these 58 combinations also had a sampling coverage of >0.85. Thus, federal-state datasets were overall less well sampled than GBIF datasets. There were also differences between taxa in how well they were sampled. Within the GBIF dataset, birds were the best-sampled taxon with 18 out of 23 cities fulfilling our criteria, followed by mammals (15 cities), lepidoptera (15 cities) and hymenoptera (14 cities) (Supplementary Appendix D). For all other taxa, less than half of the cities reached our criteria. Araneae were particularly poorly sampled and only 2 out of 18 (German states) and 4 out of 23 (GBIF) cities remained for further analysis (Table 1; Supplementary Appendix C and D, Fig. C1).

Occurrence of species within the city and in the region

For all taxa, mean species richness was higher in the region than in the city (Table 1). In general, the higher the fraction of individuals sampled within cities, the higher the proportion of species within the entire 50-km buffer that were also found within the city (Supplementary Appendix C, Fig. C2). All cities in the study comprised <4% of the buffer area, except for Hamburg (9.5%) and Berlin (11.4%); thus, the area around the buffer was much larger than the city area for all cities.

Across cities, taxa and databases, cities contained between 20.5% and 63.0% of the species of the regional species pool in

Table 1:Average species richness in cities and in the regions around the cities in Germany. Data were obtained for a radius of 50km around the city centre from GBIF (Global Biodiversity Information Facility) and from German state authorities (GS). City borders were used to delineate cities from regions. The regional species richness is the species richness outside the city borders, but inside the 50km radius. Grey shadings indicate that no data was available. To compare city and regional species richness we used paired t-tests. *-P<0.05, **-P<0.01, *-P<0.001**

Taxon	Data base	Cities with >50 observations of a taxon within city borders				Cities with >50 observations of a taxon within city borders and sample coverage >0.85			
		Mean richness cities	Mean richness region	Number of cities	Test statistic	Mean richness cities	Mean richness region	Number of cities	Test statistic
Amphibia	GBIF	11 ± 2	18 ± 1	10	t = 6.65***	11 ± 2	18 ± 1	10	t = 6.65***
	GS	12 ± 2	16 ± 2	10	t = 6.32***	12 ± 3	16 ± 3	10	t = 6.32***
Aves	GBIF	164 ± 26	275 ± 12	20	t = 9.49***	181 ± 20	277 ± 19	18	t = 13.59***
	GS	121 ± 52	195 ± 62	6	t = 4.52**	123 ± 84	191 ± 65	6	t = 4.52**
Mammalia	GBIF	31 ± 6	61 ± 7	16	t = 8.81***	32 ± 7	62 ± 13	15	t = 8.38***
	GS	31 ± 9	57 ± 7	11	t = 7.31***	32 ± 12	58 ± 12	11	t = 7.31***
Reptilia	GBIF	7 ± 2	11 ± 3	5	t = 3.93*	7 ± 2	11 ± 2	5	t = 3.93*
	GS	7 ± 1	9 ± 1	7	t = 2.93*	7 ± 1	9 ± 1	7	t = 2.93*
Coleoptera	GBIF	101 ± 28	241 ± 33	14	t = 7.29***	131 ± 47	237 ± 55	8	t = 4.3**
	GS	68 ± 68	256 ± 90	7	t = 7.41***	77 ± 77	256 ± 97	7	t = 7.41***
Diptera	GBIF	48 ± 14	87 ± 13	11	t = 5.41***	52 ± 20	91 ± 20	9	t = 4.36**
	GS								
Hemiptera	GBIF	45 ± 12	91 ± 12	13	t = 5.98***	46 ± 14	95 ± 18	12	t = 5.45***
	GS								
Hymenoptera	GBIF	43 ± 13	90 ± 10	15	t = 7.52***	46 ± 16	91 ± 18	14	t = 6.98***
	GS	57 ± 47	150 ± 175	3	t = 2.98	35 ± NA	75 ± NA	1	NA
Lepidoptera	GBIF	156 ± 56	405 ± 49	19	t = 7.13***	171 ± 62	406 ± 88	15	t = 5.87***
	GS	209 ± 109	547 ± 101	9	t = 10.62***	245 ± 127	572 ± 96	8	t = 11.29***
Orthoptera	GBIF	25 ± 7	46 ± 7	10	t = 5.69***	26 ± 7	47 ± 10	10	t = 5.69***
	GS	31 ± 4	46 ± 7	7	t = 7.4***	31 ± 4	46 ± 7	7	t = 7.4***
Araneae	GBIF	39 ± 26	136 ± 45	8	t = 4.1**	53 ± 65	109 ± 32	4	t = 2.32
	GS	37 ± 286	40 ± 133	2	t = 0.25	37 ± 286	40 ± 15	2	t = 0.25
All Taxa	GBIF	531 ± 154	1141 ± 200	20	t = 8.54***	503 ± 180	953 ± 482	17	t = 6.25***
	GS	357 ± 180	872 ± 318	11	t = 7.25***	403 ± 220	883 ± 480	9	t = 5.9***

the GBIF dataset, and between 20.4% and 69.5% in the German states dataset, with an average of 38.9 ± 7.1% (mean ± 95% confidence interval) in the GBIF dataset and 34.5 ± 12% in the German states datasets (Fig. 1a). When only taxa and cities with a sample coverage >0.85 were included, an average of 44.9 ± 7.2% of species of the regional species pool were also found in cities in the GBIF dataset, and 40.8 ± 9.6% in the German states dataset (Fig. 1b). For vertebrates, on average more than half of the species were found both in the region and in the city (Fig. 1, Supplementary Appendix E, Table E1). For arthropods, there was more variation between taxa: on average more than half of all species recorded were found both in the region and in the cities for Diptera and Orthoptera, while the average was smaller than 50% for Aranea, Hemiptera, Lepidoptera and Hymenoptera (Fig. 1; Supplementary Appendix E, Table E1).

When all cities were considered together, the percentage of species of the regional species pool that was found in at least one of the 23 cities increased to 83.2% (GBIF), and to 75.3%

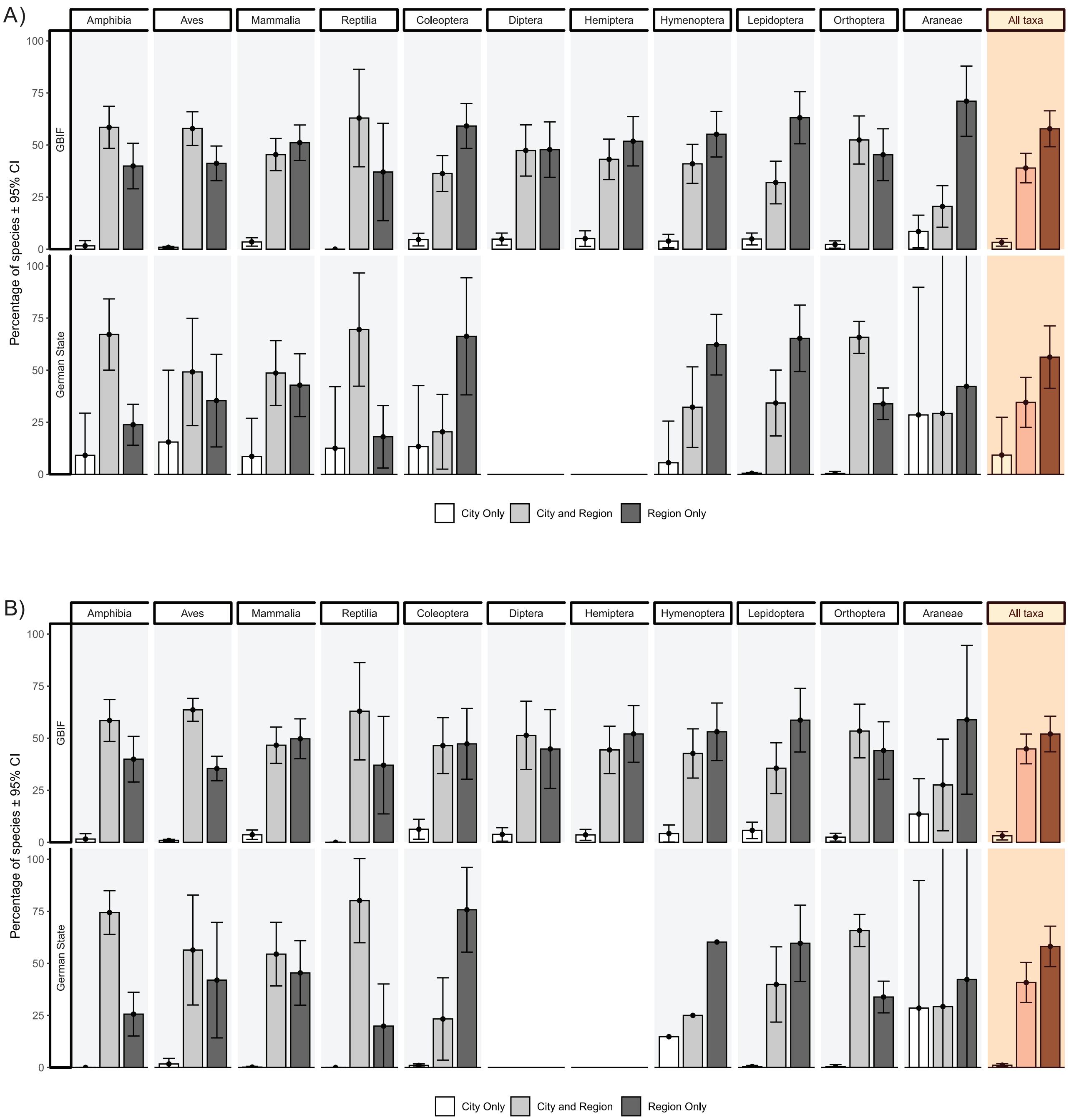

Figure 1: Occurrence of species in cities and the surrounding regions in Germany. Data were obtained for a radius of 50 km around the city centre from GBIF and from German state authorities. Administrative city borders were used to delineate cities from their surrounding region. White bars represent the proportion of species that were only found within the city borders, light grey bars represent the proportion of species that were found both within the city borders and the surrounding region, dark grey bars represent the proportion of species that were exclusively present in the surrounding region. (A) Only cities where a particular taxon had >50 observations. (B) only cities with sample coverage >0.85 for a taxon within the city. The red-shaded graphs show the proportions for all taxa combined. For the sample size (number of cities) per taxon, see Supplementary Appendix D, Table D1

(German states), when cities with >50 observations for a taxon were considered (Fig. 2a). This changed to 84.1% (GBIF) and 74.2% (German states) for cities with sample coverage >0.85 (Fig. 2b). Overall, this emphasizes that many species can occur in cities, of which many currently occur only in some of the cities.

Across cities, between 37.0% and 71.0% (GBIF) and between 18.0% and 66.3% (German states) of species only occurred in the regions around the city, with an overall average of 57.8 ± 8.6% (GBIF) and 56.3 ± 15% (German states). Applying the >0.85 sample coverage criterion changed the averages to 52 ± 8.5% (GBIF) and 58.1 ± 9.7% (German states). When data from all cities and regions were pooled, the percentage of species occurring only in regions dropped to 16.2% (GBIF) and 23.2% (German states) before coverage selection (Fig. 2a), and to 14.6% (GBIF) and 23.5% (German states) after >0.85 coverage selection (Fig. 2b).

A few species only occurred within cities. Across cities, this was the case for between 0% and 8.5% of species (GBIF) and between 0% and 28.5% of species (German states), with an overall

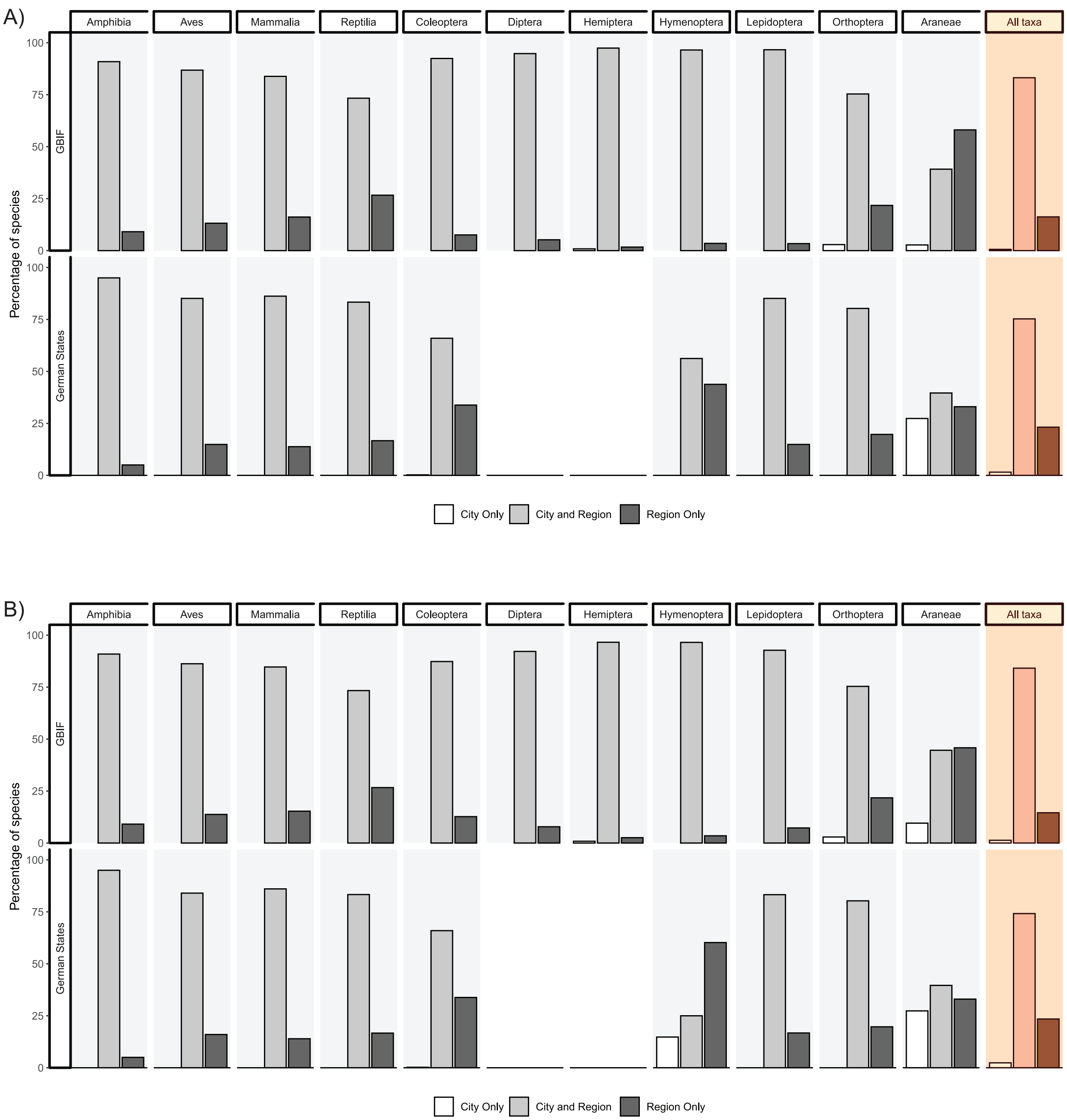

Figure 2: Occurrence of species in cities and the surrounding regions in Germany when species pools of cities and of regions were pooled across cities/regions, e.g. a species was considered to occur within cities if it occurred in any of the cities in the dataset. Data were obtained for a radius of 50 km around the city centre from GBIF and from German state authorities. City borders were used to delineate cities from their surrounding region. (A) Only cities where a particular taxon had >50 observations. (B) Only cities with sample coverage >0.85 for a taxon within the city. The red-shaded graphs represent all taxa combined

average of 3.3 ± 1.8% and 3.1 ± 2% (GBIF), and 9.2 ± 18.2% and 1.1 ± 0.8% (German states) before and after coverage selection, respectively. When all cities were pooled, only 0.6% (GBIF) and 1.5% (German states) of all the species were found exclusively within city borders before sample coverage selection, and 1.3% (GBIF) and 2.4% (German states) of all species were found exclusively within city borders after >0.85 sample coverage selection (list of these species in Supplementary Appendix F Table F1).

Discussion

In this study, we used data from public and government-owned databases for a large number of taxa to investigate the proportion of species of the regional species pool that also occurs within city borders. To standardize our approach, we used a fixed radius around the city centre and administrative borders of cities to delineate cities from regions. We compared the species richness of different taxonomic groups outside the city border (regional species pool) with their species richness inside the city. The regions outside the cities were always much larger in area than the focal city and reflected the typical cultural landscape of Germany, with different land uses such as agriculture, forestry, or nature conservation, but also other villages and towns. Our results showed that across cities and taxa on average 44.9 ± 7.2% of the species of the regional species pool occurred within city borders in the better sampled GBIF dataset (ranging from 27.5 ± 22% to 63.6 ± 5.5% between taxa). Importantly, in the same dataset, more than 80% of all of the species present in the regions around the cities were actually present in at least one of the 23 cities; between the different taxa this ranged from 39.2–97.4%. This indicates that in effect a major part of overall biodiversity can be present in cities. Our analysis also showed that the government-owned databases feature far fewer species records than GBIF and, therefore, taxa were often undersampled in these databases. There were also conspicuous differences between the different taxa. Birds were sampled well and more cities could be included in the analysis, at least in the GBIF dataset. In contrast, spiders were hardly sampled, and many cities had to be excluded from the analysis. In the remaining cities, a low proportion of spider species from the regional species pool was found, likely due to undersampling. For reptiles and amphibians, a surprisingly high fraction of species could be found in cities. Overall, our results suggest that a large part of the regional species pool is present in cities in Germany, as represented by the cities' administrative borders. Unfortunately, the data from the databases do not allow an assessment of whether the species observed within the city borders form stable urban populations, or whether there is a continuous exchange with populations in the surrounding regions. Understanding the assembly of species communities within urban borders remains an important task for urban ecology (Tzortzakaki et al. 2019).

With respect to the best-studied taxa—birds—our results are in line with previous studies that have found that in central and northern European cities, especially those at higher latitudes, a large proportion of birds of the regional species pool can occur within city borders (Ferenc et al. 2014). Here, we show that this is also true for other vertebrate taxa, i.e. amphibians, reptiles and mammals, and for the taxa within the invertebrates for which data were available, i.e. Coleoptera, Hymenoptera, Lepidoptera, Orthoptera, and Araneae in both datasets, and Diptera and Hemiptera in the GBIF dataset. Comparisons of species richness between cities and their regions are so far largely lacking for species groups other than birds. However, the few studies that exist have similar findings. For example, within mammals, settlements play an important role in bat diversity in Southern Germany (Mehr et al. 2011). Similarly, mammalian species richness has been found to be as high in developed as in natural areas in two cities and their surrounds in the eastern USA (Parsons et al. 2018). For arthropods, the occurrence in cities seems to depend on the taxa studied. For most taxa, we found that about half of all species also occurred in cities. Other studies have also found a high insect species richness in cities, for example among wild bees (Theodorou et al. 2020). Lepidoptera was the insect taxa with the on average lowest species richness shared between regions and cities in the better-sampled GBIF dataset, which is in line with a recent study, showing that Lepidoptera copes less well with urbanization (Theodorou et al. 2020). Our data also show that sampling effort differs markedly between arthropod species, with most attention directed to Hymenoptera and Lepidoptera species.

Early studies in urban ecology have often assessed species richness along a gradient of urbanization, from places outside the city via suburbs to the city centre, where the proportion of impervious surface is very high. In general, species richness is lowest at the most urban sites, i.e. the urban endpoints of the gradients that are dominated by high levels of impervious surface, whereas at intermediate levels of urbanization species, richness can be surprisingly high (McDonnell and Hahs 2008; McKinney 2008; Fenoglio et al. 2020; Callaghan et al. 2021). Therefore, when only the endpoints of the gradients are considered, the difference between the species communities living within city borders, and the communities outside the city may be exaggerated, because the city is more than just its centre. In fact, species communities outside the city differ strongly between different habitats, e.g. whether it is a forest, grassland, bog or agriculture. Similarly, the urban community is not a single community, but differences in abiotic factors, greenness and urban texture create a mosaic of habitats within city borders that allow different species to colonize different local habitats (Nilon et al. 2011). For example, local floral resources have a positive effect on hymenoptera species richness (Burdine and McCluney 2019), the availability of high-quality aquatic environments influences the presence of amphibians (Villaseñor et al. 2017), and large trees support a high bird species richness (Le Roux et al. 2018). In addition, anthropogenic food sources and the ability of species to use and compete for them may have a strong impact on urban species communities (Galbraith et al. 2015). For example, anthropogenic food sources may determine urban bird species composition especially in winter (Tryjanowski et al. 2015). Besides these local factors within the city, landscape-level factors likely also play a role (Villaseñor et al. 2017). In landscapes strongly influenced by humans such as in most temperate regions, communities are likely to change gradually across the city border, and also within the city. Therefore, what is being compared to each other is important to consider when discussing the differences between the animal communities of urban and rural areas and a continuous assessment of urbanization may be preferable.

Strikingly, when species richness was pooled across all cities, more than 80% of the species within the regional species pools were found in at least one of the cities. Our study thus emphasizes that for a collection of cities, gamma diversity of all cities combined is much higher than average city alpha diversity. This is likely due to a number of factors, including differences between cities in the diversity of the surrounding areas, and differences between cities with respect to urban texture and greenness. For example, larger, more heterogeneous cities can have a higher bird species richness than smaller cities (Callaghan et al. 2021), and differences in mammalian species richness between cities has been found to depend on the proportion of green space and housing density (Fidino et al. 2021). Global studies have found that urban biodiversity within cities increases with increasing amount of greenness, in particular patch area, and corridors have a strong positive effects on biodiversity, complemented by vegetation structure (Beninde et al. 2015). These findings emphasize the importance of planning and management for biodiversity in cities to attain biodiverse cities.

There is an ongoing debate on how to delineate regional species pools. The delineation of a region and the corresponding regional species pool is heavily dependent on what is being investigated, and the format of the available data (Cornell and Harrison 2014). The spatial delimitation of regional species pools can be based on, for example, trial and error (e.g. Bruelheide et al. 2020), biogeographical delimitations (e.g. Karger et al. 2020), map units (e.g. Ferenc et al. 2014; Fournier et al. 2019) and others (Cornell and Harrison 2014). Here, we used a fixed map unit, i.e. a 50-km radius around the city centre, from the perspective that species in the city are recruited from the region where the city occurs. The radius of 50 km was chosen as a balance between choosing an area that is large enough to represent the area from which species could immigrate into the city, to not underestimate what the species pool of the surrounding region is, and choosing an area well beyond the dispersal range of many animals. The regional species pool needs to be large enough to encompass species that may be able to disperse to cities, but that may not be able to colonize it because of habitat filters. In our case, the regional species pool included most species that were found within the respective cities (except for *Aranea*, Fig. 1, which are strongly undersampled in the region). In addition, the mean average dispersal distances of the majority of vertebrate species occurring in Germany are <50 km (e.g. for birds Paradis et al. 1998; for mammals Santini et al. 2013). Although large mammals and some larger birds may have longer dispersal distances, one also has to take into account that in highly modified landscapes such as those of Germany, actual dispersal distances are much lower due to roads and other dispersal limitations (e.g. amphibians Cayuela et al. 2019). Thus, 50 km was deemed an appropriate distance. In addition to the spatial delineation of species pools, the definition of urban species pools is not uniform across studies. Urban pools can be very inclusive (e.g. Aronson et al. 2016), as to include any and all species that exist within the city, and they can exclude introduced and/or exotic species (e.g. Ferenc et al. 2014). We considered all species that live and endure outside of direct human control in cities, therefore excluding the animals occurring in botanical gardens or zoos.

Large-scale species occurrence databases offer new opportunities for answering ecological questions, because they combine data from different sources. Data from GBIF are used increasingly in ecological research (Beck et al. 2014). Previous research has pointed out pitfalls in using GBIF data, such as its spatial bias (Beck et al. 2014). In contrast, databases administrated by government institutions are less often used by the scientific community, largely because of the difficulties in accessing the data. In Germany, the situation is complicated because each of the 16 states has their own database system and there is little homogenization among them, and basically no data exchange. In fact, some states have no database yet or are only beginning to fill them (e.g. Baden-Württemberg). More difficult still, the states have their own systems of reference names that differ from one another and from the name references of the federal agency for conservation (BfN), the conservation authority of the central government. Another challenge was that the 50-km buffer often included data from different German states that had to be combined for the analysis. Nevertheless, government databases host large datasets for taxa such as birds, bats, reptiles, amphibians and butterflies, taxa that are important in day-to-day conservation work. Our analysis shows that in order to be able to use both GBIF and government databases, careful processing of the data is mandatory. The workflow necessary to use data from GBIF and especially state databases still represents considerable work, as has been found by other authors (Zizka et al. 2020). Increased data curation, for example of coordinates, would increase the usability of the databases. Similarly, more information on the origin of data, e.g. more metadata with respect to the purpose of observation, particular circumstances (e.g. observations in a zoo) would facilitate further data use. Nevertheless, the databases are now informative enough to do ecological analyses. The workflow we applied can be used in other studies; however, it still includes a number of time-consuming steps that cannot be easily automated, such as rectifying taxonomies used in datasets.

Our study shows that for many taxa and cities, sampling coverage in public databases in Germany is already quite good. Nevertheless, many taxa are still largely undersampled, such as spiders. For other taxa, there are large differences in the sampling effort between cities, and these differences contributed to the differences in alpha diversity between cities. As this study has focused on cities in Germany, it would be interesting to conduct similar studies on other cities worldwide. Additionally, future studies could consider what factors cause differences in species richness for different taxa between cities, potentially using an SDM approach on the city level with many city-relevant environmental variables. In this study, we refrained from a comparison of alpha-diversities among cities, but with increasing completeness of the datasets such comparisons will be possible.

Conclusions

Our study emphasizes that for a large proportion of species, German cities can offer habitat. Our study thus supports the view that urban environments can be rich in biodiversity, suggesting that it is worthwhile to further understand the factors that foster urban biodiversity, and plan, design and manage cities to offer suitable habitat structures. For successful wildlife-inclusive design and management of cities, it will also be important to increase our understanding of how urban structure and urban stressors affect individuals and their success within a species, to avoid the creation of ecological traps. Our study also provides a further example of how public species occurrence data can be exploited to answer ecological questions, showing that investments in these databases, in particular provisioning of data and data curation is useful for urban ecology, and for ecology in general.

Funding

This work was funded by the Bavarian Ministry for Science and the Arts, in the program BAYKLIF—Bavarian Network for Climate Research, project BAYSICS - Bayerisches Synthese-Informations-Citizen Science Portal für Klimaforschung und Wissenschaftskommunikation - grant number 6100243, and by the German Federal Agency for Nature Conservation (BfN) grant number 351689216A.

Supplementary data

Supplementary data are available at *JUECOL* online.

Conflict of interest statement. None declared.

References

Aronson, M. F. J. et al. (2014) 'A Global Analysis of the Impacts of Urbanization on Bird and Plant Diversity Reveals Key Anthropogenic Drivers', *Proceedings of the Royal Society B: Biological Sciences*, **281**: 20133330.

—— et al. (2016) 'Hierarchical Filters Determine Community Assembly of Urban Species Pools', *Ecology*, **97**: 2952–63.

Beck, J. et al. (2014) 'Spatial Bias in the GBIF Database and Its Effect on Modeling Species' Geographic Distributions', *Ecological Informatics*, **19**: 10–5.

Beninde, J., Veith, M., and Hochkirch, A. (2015) 'Biodiversity in Cities Needs Space: A Meta-Analysis of Factors Determining Intra-Urban Biodiversity Variation', *Ecology Letters*, **18**: 581–92.

Blick, T. et al. (2016) 'Rote Liste der Spinnen', *Naturschutz und Biologische Vielfalt*, **70**: 383–510.

Bonier, F., Martin, P. R., and Wingfield, J. C. (2007) 'Urban Birds Have Broader Environmental Tolerance', *Biology Letters*, **3**: 670–3.

Bruelheide, H. et al. (2020) 'Deriving Site-Specific Species Pools from Large Databases', *Ecography*, **43**: 1215–28.

BUE. (2018). Behörde für Umwelt und Energie Hamburg: Datenbank.

Burdine, J. D., and McCluney, K. E. (2019) 'Interactive Effects of Urbanization and Local Habitat Characteristics Influence Bee Communities and Flower Visitation Rates', *Oecologia*, **190**: 715–23.

Callaghan, C. T. et al. (2021) 'How to Build a Biodiverse City: Environmental Determinants of Bird Diversity within and among 1581 Cities', *Biodiversity and Conservation*, **30**: 217–34.

Cayuela, H. et al. (2019) 'Transport Infrastructure Severely Impacts Amphibian Dispersal Regardless of Life Stage', *Scientific Reports*, **9**: 8214.

Chao, A., and Jost, L. (2012) 'Coverage-Based Rarefaction and Extrapolation: Standardizing Samples by Completeness Rather than Size', *Ecology*, **93**: 2533–47.

Chen, G. et al. (2020) 'Global Projections of Future Urban Land Expansion under Shared Socioeconomic Pathways', *Nature Communications*, **11**: 537.

Cornell, H. V., and Harrison, S. P. (2014). 'What Are Species Pools and When Are They Important?' *Annual Review of Ecology Evolution and Systematics*, **45**: 45–57.

Duchamp, J. E., and Swihart, R. K. (2008) 'Shifts in Bat Community Structure Related to Evolved Traits and Features of Human-Altered Landscapes', *Landscape Ecology*, **23**: 849–60.

Fenoglio, M. S., Rossetti, M. R., and Videla, M. (2020) 'Negative Effects of Urbanization on Terrestrial Arthropod Communities: A Meta-Analysis', *Global Ecology and Biogeography*, **29**: 1412–29.

Ferenc, M. et al. (2014) 'Are Cities Different? Patterns of Species Richness and Beta Diversity of Urban Bird Communities and Regional Species Assemblages in Europe', *Global Ecology and Biogeography*, **23**: 479–89.

Fidino, M. et al. (2021) 'Landscape-Scale Differences among Cities Alter Common Species' Responses to Urbanization', *Ecological Applications*, **31**: 1–12.

Fournier, B., Frey, D., and Moretti, M. (2019) 'The Origin of Urban Communities: From the Regional Species Pool to Community Assemblages in City', *Journal of Biogeography*, **47**: 615–29.

Galbraith, J. A. et al. (2015) 'Supplementary Feeding Restructures Urban Bird Communities', *Proceedings of the National Academy of Sciences of the United States of America*, **112**: E2648–57.

GBIF: The Global Biodiversity Information Facility. (2020). *What is GBIF?* <https://www.gbif.org/what-is-gbif>.

GBIF.org. (2020a) *GBIF Occurrence Download Amphibia*. <https://doi.org/10.15468/dl.m8ebsn>.

——. (2020b) *GBIF Occurrence Download Arachnida*. <https://doi.org/10.15468/dl.maqkjw>.

——. (2020c) *GBIF Occurrence Download Aves*. <https://doi.org/10.15468/dl.53mwq3>.

——. (2020d) *GBIF Occurrence Download Insecta*. <https://doi.org/10.15468/dl.52zxxx>.

——. (2020e) *GBIF Occurrence Download Mammalia*. <https://doi.org/10.15468/dl.gcftqj>.

——. (2020f) *GBIF Occurrence Download Reptilia*. <https://doi.org/10.15468/dl.hjm93c>.

Greenwood, J. J. D. (2007) 'Citizens, Science and Bird Conservation', *Journal of Ornithology*, **148**: 77–S124.

Grimm, N. B. et al. (2008) 'Global Change and the Ecology of Cities', *Science (New York, N.Y.)*, **319**: 756–60.

Grüneberg, C. et al. (2015) 'Rote Liste Der Brutvögel Deutschlands', *Berichte Zum Vogelschutz*, **52**: 19–67.

Guetté, A. et al. (2017) 'Measuring the Synanthropy of Species and Communities to Monitor the Effects of Urbanization on Biodiversity', *Ecological Indicators*, **79**: 139–54.

Haack, N. et al. (2021) 'Patterns of Richness across Forest Beetle Communities—a Methodological Comparison of Observed and Estimated Species Numbers', *Ecology and Evolution*, **11**: 626–35.

Harris, S. J. et al. (2020). 'The Breeding Bird Survey', 726.

Hsieh, T. C., Ma, K. H., and Chao, A. (2016) 'INEXT: An R Package for Rarefaction and Extrapolation of Species Diversity (Hill Numbers)', *Methods in Ecology and Evolution*, **7**: 1451–6.

Ibáñez-Álamo, J. D. et al. (2017) 'Global Loss of Avian Evolutionary Uniqueness in Urban Areas', *Global Change Biology*, **23**: 2990–8.

iNaturalist. (n.d.). *INaturalist*. <https://www.inaturalist.org>.

Jung, K., and Threlfall, C. G. (2018) 'Trait-Dependent Tolerance of Bats to Urbanization: A Global Meta-Analysis', *Proceedings of the Royal Society B: Biological Sciences*, **285**: 20181222.

Karger, D. N. et al. (2020) 'Disentangling the Drivers of Local Species Richness Using Probabilistic Species Pools', *Journal of Biogeography*, **47**: 879–89.

Knop, E. (2016) 'Biotic Homogenization of Three Insect Groups Due to Urbanization', *Global Change Biology*, **22**: 228–36.

Kühnel, K.-D. et al. (2009) 'Rote Liste und Gesamtartenliste der Lurche (Amphibia) Deutschlands—Stand Dezember 2008', *Naturschutz und Biologische Vielfalt*, **70**: 259–88.

LANUV NRW. (2018) Datenlizenz Deutschland – Namensnennung – Version 2.0. <www.govdata.de/dl-de/by-2-0>. LINFOS Landschaftsinformationssammlung – Planungsrelevante Arten. Land NRW.

LAU. (2018) Landesamt für Umweltschutz Sachsen-Anhalt: Datenbank.

Le Roux, D. S. et al. (2018) 'The Value of Scattered Trees for Wildlife: Contrasting Effects of Landscape Context and Tree Size', *Diversity and Distributions*, **24**: 69–81.

Leveau, L. M. (2021) 'Temporal Persistence of Taxonomic and Functional Composition in Bird Communities of Urban Areas: An Evaluation after a 6-Year Gap in Data Collection', *Urban Ecosystems*, **25**: 9–20

LfU Bayern. (2018) Bayerisches Landesamt für Umwelt.

LfU Rheinland-Pfalz. (2018) Landesamt für Umwelt—Rheinland-Pfalz: Datenbank.

LfULG. (2018) Darstelliung unter Verwendung von Daten aus der Zentralen Artdatenbank Sachsen LfULG vom December 2018.

LLUR. (2018) Landesamt für Landwirtschaft, Umwelt und ländliche Räume- Schleswig Holstein: Datenbank.

LUBW. (2018) Landesanstalt für Umwelt Baden-Württemberg: Datenbank.

LUNG. (2018) Landesamt für Umwelt, Naturschutz und Geologie Mecklenburg-Vorpommern: Datenbank.

LVGL. (2018) Landesamt für Vermessung, Geoinformation und Landentwicklung- Saarland: Datenbank.

Maas, S., Detzel, P., and Staudt, A. (2011) 'Rote Liste und Gesamtartenliste der Heuschrecken (Saltatoria) Deutschlands', *Naturschutz Und Biologische Vielfalt*, **70**: 577–606.

McDonnell, M. J. and Hahs, A. K. (2008) 'The Use of Gradient Analysis Studies in Advancing Our Understanding of the Ecology of Urbanizing Landscapes: Current Status and Future Directions', *Landscape Ecology*, **23**: 1143–55.

McKinney, M. L. (2006) 'Urbanization as a Major Cause of Biotic Homogenization', *Biological Conservation*, **127**: 247–60.

—— (2008) 'Effects of Urbanization on Species Richness: A Review of Plants and Animals', *Urban Ecosystems*, **11**: 161–76.

Mehr, M. et al. (2011) 'Land Use is More Important than Climate for Species Richness and Composition of Bat Assemblages on a Regional Scale', *Mammalian Biology*, **76**: 451–60.

Meinig, H. et al. (2020) 'Rote Liste und Gesamtartenliste der Säugetiere (Mammalia) Deutschlands', *Naturschutz und Biologische Vielfalt*, Vol. **170, pp. 74**.

Merckx, T. et al. (2018) 'Body-Size Shifts in Aquatic and Terrestrial Urban Communities', *Nature*, **558**: 113–6.

Morelli, F. et al. (2016) 'Evidence of Evolutionary Homogenization of Bird Communities in Urban Environments across Europe', *Global Ecology and Biogeography*, **25**: 1284–93.

Nilon, C. H., Warren, P. S., and Wolf, J. (2011) 'Baltimore Birdscape Study: Identifying Habitat and Land-Cover Variables for an Urban Bird-Monitoring Project', *Urban Habitats*, **6**: 1–13.

NKWKN. (2018) Tierarten-Erfassungsprogramm der Fachbehörde für Naturschutz im Niedersächsischen Landesbetrieb für Wasserwirtschaft, Küsten- und Naturschutz.

OpenStreetMap Contributors. (2021) *OpenStreetMap*. <www.openstreetmap.org>.

Paradis, E. et al. (1998) 'Patterns of Natal and Breeding Dispersal in Birds', *Journal of Animal Ecology*, **67**: 518–36.

Parsons, A. W. et al. (2018) 'Mammal Communities Are Larger and More Diverse in Moderately Developed Areas', *eLife*, **7**: 1–13.

Pedersen, T. L. (2019). *patchwork: The Composer of Plots* (1.0.0).

Piano, E. et al. (2020) 'Urbanization Drives Cross-Taxon Declines in Abundance and Diversity at Multiple Spatial Scales', *Global Change Biology*, **26**: 1196–211.

QGIS Development Team. (2021). *QGIS Geographic Information System*. QGIS Association.

R Development Core Team. (2008). *R: A Language and Environment for Statistical Computing. (3.4.1)*. R Foundation for Statistical Computing, Vienna, Austria.

Rote-Liste-Gremium Amphibien und Reptilien. (2020) 'Rote Liste und Gesamtartenliste der Reptilien (Reptilia) Deutschlands', *Naturschutz und Biologische Vielfalt*, **170, pp. 65**.https://doi.org/10.19213/972173/.

Rstudio Team. (2015). *RStudio: Integrated Development for R (3.3.3)*. Rstudio, Inc.

Saari, S. et al. (2016) 'Urbanization is Not Associated with Increased Abundance or Decreased Richness of Terrestrial Animals—Dissecting the Literature through Meta-Analysis', *Urban Ecosystems*, **19**: 1251–64.

Santini, L. et al. (2013) 'Ecological Correlates of Dispersal Distance in Terrestrial Mammals', *Hystrix, The Italian Journal of Mammalogy*, **24**:181–6.

Schwarz, J., and Flade, M. (2000) 'Ergebnisse des DDA-Monitoringprogramms Teil I: Bestandsänderungen von Vogelarten der Siedlungen seit 1989', *Vogelwelt*, **121**: 87–106.

Stiftung Naturschutz Berlin. (2018) Koordinierungstelle Fauna. Stiftung Naturschutz Berlin, im Auftrag der Senatsverwaltung für Stadtentwicklung und Umwelt: Fauna Berlin—Nachweise und Erfassung (Stand 1901-14.09.2018)—Export digitaler Originaldaten der Faschale Fauna.

Sullivan, B. et al. (2009) 'EBird: A Citizen-Based Bird Observation Network in the Biological Sciences eBird: A Citizen-Based Bird Observation Network in the Biological Sciences', *Biological Conservation*, **142**: 2282–92.

Theodorou, P. et al. (2020) 'Urban Areas as Hotspots for Bees and Pollination but Not a Panacea for All Insects', *Nature Communications*, **11**: 1–13.

Threlfall, C. G. et al. (2016) 'Approaches to Urban Vegetation Management and the Impacts on Urban Bird and Bat Assemblages', *Landscape and Urban Planning*, **153**: 28–39.

TLUG. (2018) Thüringer Landesanstalt für Umwelt und Geologie: Referat Zoologischer Artenschutz, Vogelschutzwarte Seebach.

Trimaille, E. (2020). *QuickOSM: QGIS plugin to fetch OSM data with the Overpass API* (1.14.3).

Tryjanowski, P. et al. (2015) 'Winter Bird Assemblages in Rural and Urban Environments: A National Survey', *PLOS One*, **10**: e0130299.

Turrini, T., and Knop, E. (2015) 'A Landscape Ecology Approach Identifies Important Drivers of Urban Biodiversity', *Global Change Biology*, **21**: 1652–67.

Tzortzakaki, O. et al. (2019) 'Butterfly Diversity along the Urbanization Gradient in a Densely-Built Mediterranean City: Land Cover is More Decisive than Resources in Structuring Communities', *Landscape and Urban Planning*, **183**: 79–87.

University of California Berkely. (2018). *Global Administrative Areas (GADM)*. <www.gadm.org>.

Villaseñor, N. R. et al. (2017) 'The Relative Importance of Aquatic and Terrestrial Variables for Frogs in an Urbanizing Landscape: Key Insights for Sustainable Urban Development', *Landscape and Urban Planning*, **157**: 26–35.

VSWFFM. (2018) Dartstellung auf der Grundlage von Daten der Staatlichen Vogelschutzwarte für Hessen, Rheinland-Pfalz und Saarland.

Weisser, W. W., and Hauck, T. E. (2017) 'ANIMAL-AIDED DESIGN – Using a Species' Life-Cycle to Improve Open Space Planning and Conservation in Cities and Elsewhere', BioRxiv, **150359**.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.

Wood, S. N. (2021). *Mixed GAM Computation Vehicle with Automatic Smoothness Estimation* (1.8-34).

Zizka, A. et al. (2019) 'CoordinateCleaner: Standardized Cleaning of Occurrence Records from Biological Collection Databases', *Methods in Ecology and Evolution*, **10**: 744–51.

—— et al. (2020) 'No One-Size-Fits-All Solution to Clean GBIF', *PeerJ*, **8**: e9916.