Effects of publication bias on conservation planning

Hickisch, R.^{1*}, Hodgetts, T.¹, Johnson, P.J.¹, Sillero-Zubiri, C.^{1,2}, Tockner, K.^{3,4,5}, Macdonald, D.W.¹

1. Wildlife Conservation Research Unit (WildCRU), Zoology, University of Oxford, The Recanati-Kaplan Centre, Tubney House, Tubney OX13 5QL, United Kingdom
2. Born Free Foundation, Horsham RH12 4QP, United Kingdom
3. Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB), 12587 Berlin, Germany
4. Austrian Science Fund (FWF), Sensengasse 1, 1090 Vienna, Austria
5. Institute of Biology, Freie Universität Berlin, Altensteinstraße 6, 14195 Berlin, Germany

[*email raffaelhickisch@gmail.com](mailto:raffaelhickisch@gmail.com)

Running head: Publication Bias

Keywords: accessibility, geographic bias, maps, publication density, sampling effort

Abstract

Conservation planners need reliable information on spatial patterns of biodiversity. However, existing data sets are skewed because some ecosystems, taxa, and locations are underrepresented. We determined how many articles have been published in recent decades on the biodiversity of different

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/cobi.13326](https://doi.org/10.1111/cobi.13326).

This article is protected by copyright. All rights reserved.

countries and their constituent provinces. We searched the Web of Science catalogues SCI and SSCI for biodiversity-related articles published from 1993 to 2016 that included country and province names. We combined data on research publication frequency with other provincial-scale factors hypothesized to affect the likelihood of research activity (i.e., economic development, human presence, infrastructure, and remoteness). Areas that appeared understudied relative to the biodiversity expected based on site climate likely have been inaccessible to researchers for reasons, notably armed conflict. Geographic publication bias is of most concern in the most remote areas of sub-Saharan Africa and South America. Our provincial-scale model may help compensate for publication biases in conservation planning by revealing the spatial extent of research needs and the low cost of redoing this analysis annually.

Introduction:

Repositories, such as the IUCN Red List and the Global Biodiversity Information Facility (GBIF), provide data that supports conservation planning; however, knowledge of the abundance and distribution of biodiversity remains patchy. The global number of species is unknown, and spatial distribution of species that have been described is often uncertain (knowledge deficits termed Linnaean and Wallacean shortfalls, respectively) (Brito 2010). Hence, there are continuing calls to improve biodiversity information (Jetz et al. 2012). Indeed, ‘maps of ignorance’ quantifying knowledge deficits would be useful (Rocchini et al. 2011).

The spatial patchiness of existing biodiversity data sets reflects historic and contemporary research efforts. Hence, these data sets incorporate biases and omissions (e.g., Boakes et al. 2010). Vertebrates and forests, for example, are favored (Fazey et al. 2005), and research biases within taxonomic groups

exist. For example, field research is biased toward large species (Brodie, 2009), and body size, range size, and diet explain uneven patterns of research efforts across carnivores (Brooke et al. 2014).

These uneven spatial patterns of research effort occur for a variety of reasons. Accessibility is highly influential. For example, sampling locations of passerine birds in sub-Saharan Africa are clustered near roads, along rivers, and around urban areas (Reddy & Dávalos 2003). The practicalities of working and living conditions are important. Socioeconomic conditions (e.g. levels of affluence, infrastructure, language, security) privilege specific areas for research and thus influence spatial sampling patterns (Martin et al. 2012; Amano & Sutherland, 2013; Meyer et al. 2016). For example, research efforts on coral reefs are positively correlated with an area's per capita gross domestic product (Fisher et al. 2011). Research effort also tends to be biased against areas that have not been studied previously. Although this could be explained by a continuing need to tackle threats to the biodiversity in these areas, it may also be due to the practical and infrastructural benefits provided by biological field stations and established monitoring sites (Tydecks et al. 2016). Thus, protected areas (PAs) are overrepresented, as are sites in temperate deciduous woodlands and wealthy countries (Martin et al. 2012). The wealth effect also influences the impact of research that gets published. Citation rates for conservation research increase as a country's wealth, standard of governance, human population, and conservation need increase (Meijard et al., 2015). Whatever the causes, these biases can lead to a self-reinforcing cycle whereby "funding begets biodiversity" (Ahrends et al. 2011).

Geographic sampling biases in biodiversity research have implications for macroecological modeling (Yang et al. 2013), conservation prioritization (De Ornellas et al. 2011), and governance audits, such

as those of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (or IPBES). Although modelers have ways to correct for sampling biases in species-range modeling (Kramer Schadt et al. 2013; Engemann et al. 2015) and in systematic conservation planning tools (Rondinini et al. 2006), such remedies cannot always address the persistent undersampling of biodiversity (Wilson et al. 2016). For example, there is little spatial congruence in the occurrence of threatened bird and mammal species at spatial scales relevant to conservation planning and congruence is particularly low for the most endangered species (Grenyer et al. 2006). Redirecting effort from well-sampled toward undersampled locations may improve the efficiency of conservation actions (Aranda et al., 2011). Under some plausible assumptions, if the aim of a study is to describe the spatial distribution of species richness in a given area, then surveying undersampled locations is more effective than surveying areas known to have high species richness (Sastre & Lobo 2009), and focusing on poorly surveyed areas maximizes benefits of surveys for conservation planning (De Ornellas et al. 20011).

We developed a global, provincial-scale map of biodiversity publication density to identify areas where low accessibility is associated with research deficits. We define *provinces* as subnational administrative units, as recognised in the Global Administrative Areas database (GADM) (GADM 2017), covering differently named areas (e.g., provinces, departments, counties) of various sizes. These include, for example, vast federal territories in Australia (e.g., Northern Territory) and small provinces such as cantons in Switzerland (e.g., Ticino).

These provinces are unrelated to the biogeographic meaning of *province*.

For each province, we mapped the number of biodiversity research publications by area based on information gathered from the Web of Science catalogues SCI (Science Citation Index) and SSCI (Social Science Citation Index). We considered the map a surrogate for sampling effort. Lobo (2008) used a similar approach. We modeled the associations between factors affecting accessibility (i.e., remoteness, infrastructure, and development) and the number of publications reporting on biodiversity. We identified provinces where, in addition to a likely shortfall in publications linked to accessibility, a human impact on biodiversity was also likely.

Methods

Creating an index of biodiversity publications

To compare how many articles were published from 1993 to 2016 on the biodiversity of different countries and their constituent provinces, we queried the Web of Science (WoS) catalogs Science Citation Index and Social Science Citation Index for publications labeled in the Biodiversity & Conservation category (covering publications appearing in any journal listed under that category; not to be confused with the journal of the same name). We queried WoS by topics (meaning a combination of title, abstract, author, and “keywords plus,” see https://images.webofknowledge.com/images/help/WOS/hs_topic.html). We searched for country and each province name in multiple languages (details in the Supporting Information) because a substantial proportion of global biodiversity research is not published in English (35.6% in 2014 [Amano et al. 2016]). We limited our search to 1993-2016 – starting with the UN Earth Summit--to capture a period of sustained global research in biodiversity. Many more biodiversity publications were identifiable to country level ($n = 1,229,636$) than to province level ($n = 54,144$). This index could

in the future be refined with more comprehensive publications repositories and improved metadata (e.g., with WoS storing the geographic scope of scientific publications).

Imputing missing values

Some provinces had no results for the biodiversity and conservation WoS category (Supporting Information). We assumed this was because biodiversity research in these areas was attributed only at a national level, rather than assuming such research did not exist (i.e., the unknown value is the province-level number of biodiversity publications). To factor in this assumption, we created an index based on imputed values for the number of biodiversity and conservation publications per province based on the national-level ratio of biodiversity research to overall research within a country. Assuming this proportion (biodiversity research: all topics research) was stable for a country, we multiplied the province's overall total number of publications (all topics) by this proportion ($n=44,858$ publications on provincial level). If a country had a total of N publications and C conservation publications, and x publications for a province, the imputed number of biodiversity publications (INP) equals $(C/N)*x$. To test whether this assumption was reasonable, we fitted a linear model for the subset of provinces where the number of publications was known and used imputed number of biodiversity publications as the response and observed number of conservation publications as the predictor (details in Supporting Information). We refer to this index as imputed number of biodiversity publications.

Queries were performed using the R package *rwos* (Barnier, 2017) and the WoS Lite API (application programming interface) (Reuters, 2017). Country and province name were from the GADM data set (GADM.org, 2017) and the R package *countrycode* (Arel-Bundock, 2017). Results were filtered in the different WoS domains (all, no constraint; cons, biodiversity and

conservation as single category). We mapped the imputed number of biodiversity publications as density (per square kilometer) and applied a log scale color viridis band in the R language version 3.1 and QGIS version 2.18 software (scripts in Supporting Information).

Generating proxy measures of factors hypothesized to affect number of publications

We hypothesized that the number of biodiversity and conservation articles written about a province would be affected by its expected level of biodiversity, extent of human presence, remoteness and accessibility for researchers, and level of socioeconomic development. We selected metrics available in similar detail and precision worldwide and unlikely to be affected by the type of sampling bias we sought to demonstrate.

We used ecological conditions as a proxy for the level of biodiversity in a province. For the index, we combined mean precipitation and temperature data derived from Worldclim models (Worldclim, 2017; Fick & Hijmans, 2017). These data sets are highly detailed and verifiability given that satellite sensor data are used as covariates and temperature and precipitation are interpolated from weather stations on the ground. Selecting these data as proxies for biodiversity is based on the latitudinal gradient theory (Gaston, 2000), which suggests biodiversity is high where humidity, temperature, and net primary productivity are high. We adopted this approach as an overall proxy despite well-known caveats (e.g., poor predictor of species' distributions for some taxonomic groups) because it was preferable to other options (e.g., IUCN Red List, GBIF repository, or other vertebrate maps [Mannion, 2014]); other measures are more likely to be affected by structural biases relating to sampling effort (particularly in Central Africa); and this bias is exactly the type we aimed to estimate.

We measured human presence based on the NASA nighttime light data set corrected for gas flares (NASA, 2017a) and weekly fire clusters of the NASA VIIRS close to real time data set (spatial

resolution minimum of 750 m) (NASA, 2017b). The former is an acknowledged measure of economic activity. The latter (publicly available only since January 2016) has not been used previously in this type of analysis; however, it has been used in some nature reserves (e.g., Garamba National Park) to orchestrate patrolling effort. We prefer the VIIRS fire data over other measures of human presence, such as the global terrestrial human footprint (Venter, 2016), because the footprint concept focuses largely on built infrastructure, which may be a poor indicator of the presence of people in remote regions and areas of civil war who rely on natural resources to survive. As compound measures, footprint-style indices are also difficult to interpret. We suggest such people are often far from built infrastructure, sometimes occupying inaccessible places (e.g., Joseph Kony's Lord Resistance Army). Because humans use fires for cooking, land clearing, preparing pastures for livestock, and clearing areas when hunting elephant and to ease walking across savannah, we created an associated index that captures such activities.

Within data on fire clusters, we corrected for large bushfires by spatially clustering reported fire points by week. Weekly fire clusters may be a better indicator than simple fire point counts because these remove the effects of large bush fires that can ignite naturally or result from a one-off human action (Supporting Information). These fire indices complement those quantifying human presence insofar as the latter (including the NASA nighttime light data set) fail to capture the impact of nomadic agropastoral communities. We therefore calculated a world road area index (derived from Ibisch et al. 2016) by subtracting roadless area from overall area for each province as an approximate index of built infrastructure. We tested how well these indices predicted fire clusters by fitting a statistical model with the fire clusters as a response and the predictors night light, road area, and World Database of Protected Areas category I and II PAs (Supporting Information).

We also developed a measure of relative accessibility of provinces for biodiversity researchers. Our remoteness index was based on route-planning results for mean travel time from the province polygon center point to capitals, airports, and seaports (Giraud, 2017) weighted with the mean province value of an existing accessibility raster data set (Nelson 2008). We used this combined index because travel time to major cities (Nelson, 2008), which at the time of analysis was the highest resolution dataset available, did not account for logistical challenges associated with research, where permits often need to be obtained in the capital city, vehicles and spare parts shipped from far away, and international travelers flown in from international airports for immigration and customs clearance. To establish the plausibility of the remoteness index, we fitted a statistical model with the index as a response and night-light intensity, road area, and province area as predictors (Supporting Information). This indicator can be improved by better vector data sets and route planning algorithms.

We used mean life expectancy at birth (1960-2010) as a measure of human development and living conditions in a province (World Bank Open Data, 2017) because these data are the most traditional and least correlated with other selected indices. It reflects available health-related infrastructure and level of societal conflict. We tested the Human Development Index (World Bank Open Data, 2017) as alternative indicator, but rejected it because it is relatively complex compared with life expectancy and it has limited weight in principal components analysis (PCA).

We ran a linear mixed-effects model (lmer with R lme4 package [Bates et al. 2015]) with imputed number of biodiversity publications as the response and variables for remoteness and accessibility as predictors corrected for clustering of provinces within countries by treating country as a random effect. Despite the evident circularity, we additionally incorporated PAs (IUCN & UNEP-WCMC 2017) and IUCN terrestrial mammal ranges (IUCN 2017) as a validity check against conventional indicators (Supporting Information). Terrestrial mammal ranges were selected because they appeared

to have the most complete and refined data set in the IUCN Red List data set (fungi and amphibians had larger data deficiencies).

Testing proxy measures as predictors of number of publications

Given the numerous the potential drivers affecting both accessibility and biodiversity, we used PCA to reduce the data to a small set of orthogonal axes summarising input variables. We cut all geospatial data to the extent of the province shape and calculated mean values (for raster data), size, and counts for all vector data sets and added nonspatial attributes by join via the country name. We log transformed variables with large variance and high skew (number of biodiversity publications, road area, fire clusters, and remoteness). These data are stored on file per GADM province record (details in Supporting Information). We used these data to conduct a PCA (in the `prcomp` function in R), correcting for province size by inclusion, and tested whether the indices suitably characterized provincial data. We removed 53 of 204 countries and 184 of 3,417 provinces prior to the PCA because of missing data. In most cases, missing data were related to life expectancy and applied to small island groups. These are often not listed separately in the relevant World Bank data sets. We plotted the location of provinces as a function of PC1 and PC2 scores with the `ggfortify` R package (Tang et al. 2016). We identified quadrants in the PCA plot with the Cartesian origin at scores of zero for both PC1 and PC2.

Identification of priority provinces for research and conservation

We identified priority provinces for conservation research and action as those with remoteness and accessibility most strongly linked to the number of biodiversity publications and where climate (and latitude) indicated biodiversity was likely high. We did so by summing the absolute axis scores of

PC1 and PC2 for those provinces in the quadrant most characterised by remoteness and large number of fire clusters.

Results

The highest density of publications (Fig. 1) were in affluent countries (e.g., Canada, Iceland), countries with biodiversity hotspots (e.g. Brazil, Kenya), and affluent and biodiverse countries (e.g., Australia, Costa Rica, South Africa, United States). As expected from the latitudinal biodiversity gradient theory, high-latitude areas were associated with relatively fewer biodiversity publications. The number of biodiversity publications was low for some areas close to the equator, such as Central Africa.

There was a positive and significant correlation between the number of biodiversity and conservation publications at province level and imputed number of publications ($R^2 = 98.4\%$). Based on the assumption that this number is proportional to the size of the province, the imputed number predicted almost all the variation in the real number.

The remoteness index increased as area of a province increased and decreased as night light and road density increased (total $R^2 = 45.0\%$) (Supporting Information). The fire-cluster index decreased as area electrified increased (night light intensity) and as protected area increased (total $R^2 = 61.0\%$).

The number of imputed publications of a province correlated most strongly with life expectancy, night-light intensity, and log-transformed road area (Table 2). The number of publications associated with a province increased as human-development scores and road area increased (Table 2). The

number of publications and number of PAs were positively correlated in the respective province (Table 2), as was with the number of terrestrial mammal species in a province according to IUCN distribution range maps. Given the large sample sizes, we considered effect sizes and estimates of precision rather than p values in interpreting the outputs of statistical models (main effects Fig. 2).

The PCA (Fig. 3) illustrated the patterns of intercorrelation among the input variables, and showed where research deficits (low numbers of publications) were likely to occur. The first 2 components together explained 54.1% of the variance in input variables (Supporting Information). Visually, four semiarbitrary groups of provinces were defined, roughly corresponding to the quadrants of the PCA (A – D) (Fig. 3). Observations farthest from the origin in quadrant A tended to be provinces with high levels of human development (i.e., high life expectancy) and high economic activity (night light). These provinces exhibited relatively high numbers of biodiversity publications and low temperature and low rainfall relative to provinces distant from the origin in quadrants C and D. Provinces in the outer region of quadrant B tended to have high numbers of publications, high road density, and large area and, again, low temperature relative to provinces in quadrants C and D. Quadrant C included primarily tropical provinces, with higher temperatures and precipitation, fewer publications (relative to quadrants a and b), and lower degrees of remoteness than those in quadrants B and D. Accessibility may be of less relative importance in provinces in quadrant C. Quadrant D contained provinces that had few publications but high levels of remoteness (long travel times to major cities, capitals, seaports, and airports) and distinct human presence. These provinces were in relatively less affluent regions with more fire clusters than road area and night light.

The 100 provinces of the highest research and conservation priority were largely within 15° latitude of the equator (Fig. 3), where the accessibility bias in geographic biodiversity reporting is likely most influential (quadrant D; detailed list in Supporting Information).

Discussion

We identified provinces with marked differences between the observed density of publications concerning biodiversity and conservation and that expected from their climate and latitude (Gaston 2000; Mannion et al. 2014). The discrepancies are most likely explained by accessibility. By integrating a large array of available data sets, our method may provide a tool to help optimize biodiversity monitoring and research efforts regionally and globally. We consider our provincial-scale model a first step toward correcting of present sampling bias – particularly because this may guide conservation researchers at the onset of their careers.

Most biological field stations are close to PAs, many of which are protected by virtue of their high biodiversity, which creates a positive accessibility bias (Tydecks et al. 2016). Funding opportunities are likely higher for these areas compared with unprotected biodiversity hotspots (Ahrends et al. 2011). The positive correlation between number of biodiversity publications and the distribution ranges of terrestrial mammals may reflect appropriate targeting. Nevertheless, species' distribution range maps are frequently the outcome of biased reporting efforts.

Case studies from Central Africa illustrate how understanding of spatial patterns in biodiversity is likely biased. For example, 7 of 17 provinces in the Central African Republic (CAR) were among the

100 underreported areas with accessibility issues. These included areas well known to R.H., who has worked in the region since 2012, such as the Mboumo and Haute-Kotto provinces, where the Chinko Nature Reserve was established in 2013. This reserve covers 19,840 km² and is one of the largest managed PAs in the region. Its inaccessibility results from continued violence and remoteness. The resource-rich provinces of CAR and adjacent countries have experienced cycles of conflict and civil war for 3 decades (Collier, 2007). Many rebel groups, including the Lord's Resistance Army, seek refuge in remote, ungoverned provinces likely to be rich in wildlife (Ondoua 2017). Given these accessibility problems, the current state of biodiversity reporting in the region is far from adequate, in our experience.

Using fire clusters as a complementary indicator of human influence revealed an additional category of understudied provinces in quadrant D (Fig. 3), which are likely to be of particular concern for conservation. These provinces are expected to exhibit a high level of threat to biodiversity. However, road infrastructure is poor, and security may be limited; therefore, knowledge about biodiversity may also be limited (Reddy & Dávalos 2003).

Priority provinces exhibiting distinct deficits in biodiversity-related research, primarily due to low accessibility, shared common attributes (Fig. 4). While these provinces are not necessarily rich in biodiversity, their latitude and climatic conditions suggest that high biodiversity is probable. However, they were located in remote areas, had low human-development index scores, and contained poor infrastructure. Most of these provinces were in Central Africa, central Western Africa, Bolivia, Amazonian (Peru and lowland areas of northeastern Brazil), and Andaman and Nicobar Islands (India) (Fig. 4).

Provinces that had few publications and low human presence (based on fire clusters) may also exhibit important knowledge deficits but are likely to have less threatened biodiversity. These provinces tended to be in Amazonian Brazil, Caribbean Coast of Nicaragua, Sipaliwini in Suriname, most of Gabon, and South Kalimantan and Sumatera Selatan (Indonesia).

If poorly researched provinces were randomly distributed, then adjacent provinces could be deemed representative. However, many inaccessible and underresearched provinces have faced serious geopolitical challenges in recent decades, including interstate and civil wars, rebel insurgencies, piracy, and drug wars. The frequent occurrence of armed conflicts in areas of high biodiversity means fine-scale assessments of conflict risk are desirable for efficient PA planning, and additional investment is often required (Hammill et al. 2016). However, poor security often hinders funding and field research (Amano & Sutherland 2013). For example, violent conflict can lead to the destruction of habitats, bushmeat poaching by combatants and displaced communities, poaching for illegal trade, persistent social disruption, and lack of conservation capacity (Gaynor et al. 2016; Conteh et al. 2017; Daskin & Pringle 2018).

Ignorance of the scale of processes like these is likely to severely compromise conservation planning. For example, the Chinko Basin formerly contained thousands of elephants (Douglas-Hamilton 1987). The population has since crashed to a few hundred individuals (Hickisch & Aebischer 2013) because of ivory poaching. Since 2000, poaching has also caused population declines in Garamba National Park in the Democratic Republic of Congo (Canby 2016) and in Zakouma National Park in Chad (Fay 2007). These declines occurred in inaccessible provinces for which there were few publications.

The neglect of threats may be widespread, although more research is needed in this area. There are examples of less accessible provinces in Central Africa, where significant changes in the practices of pastoral communities are occurring (Economist 2017). These changes have as yet been largely unrecognised by conservation researchers. Social changes, the legacy of violence, and economic factors are leading to larger domestic stock herd sizes, and climatic factors combined with increases in vaccine availability are allowing movement of these herds into areas formerly unsuitable for grazing (e.g., tsetse fly habitats). Changing practices of transhumance, larger herd sizes, widespread availability of automatic weapons, and highly potent poisons threaten wildlife in multiple ways, from bushmeat hunting to eradication of top predators (Bouché et al. 2010; Brugière et al. 2015). Although the movements of large herds can be remotely monitored (e.g., fire-cluster data), the subsequent impacts on wildlife (especially terrestrial mammals and birds) are more difficult to assess remotely, particularly if dense vegetation appears largely intact. The lack of adequate data on the response of wildlife to these changes is important in its own right.

Although the priority provinces largely fell outside biodiversity hotspots (Myers et al. 2000), they were almost all adjacent or connected to areas of high conservation concern. We acknowledge the difficulty of working in inaccessible areas and areas affected by conflict, but suggest greater attention be afforded to the implications of the bias that results. Some conservation practitioners and nongovernmental organizations have extended their focus to these regions, at least on the African continent. For example, the African Parks Network (<https://www.africanparks.org/>) has expanded its portfolio of private-public partnerships in Central Africa. The efficiency of these and related initiatives in terms of the biodiversity gains they achieve for a given investment would benefit from considering the patterns we found.

Acknowledgments

We thank A. Dickman, E. Dröge, E. Macdonald, H. Bauer, A. Siddig, and A. Ellison for their inputs. M. Strimas-Mackey helped with mapping in R and A. Ruete and D. Williams provided information on literature and data sets. We thank the audience in the WildCrU colloquium (Oxford 2017) and E.J. Milner-Gulland for critical questions and helpful remarks. The results of T. Aebischer's fabulous fieldwork in Central African Republic laid the foundations for this work. This work was begun while R.H. was a student on the Recanati-Kaplan Centre Postgraduate Diploma in International Wildlife Conservation Practice delivered by WildCRU.

Literature cited

Ahrends A et al. 2011. Funding begets biodiversity: Funding begets biodiversity. *Diversity and Distributions* **17**:191–200.

Amano T, Sutherland WJ. 2013. Four barriers to the global understanding of biodiversity conservation: wealth, language, geographical location and security. *Proceedings of the Royal Society B: Biological Sciences* **280**:20122649–20122649.

Amano T, González-Varo JP, Sutherland WJ. 2016. Languages are still a major barrier to global science. *PLoS Biology* **14** DOI: 10.1371/journal.pbio.2000933.

Aranda SC, Gabriel R, Borges PAV, de Azevedo EB, Lobo JM. 2011. Designing a survey protocol to overcome the Wallacean shortfall: a working guide using bryophyte distribution data on Terceira Island (Azores). *The Bryologist* **114**:611–624.

Arel-Bundock, V. 2017. countrycode R package: convert country names and country codes. Available from <https://cran.r-project.org/web/packages/countrycode/index.html> (accessed September 2017).

This article is protected by copyright. All rights reserved.

Barnier, J. 2017. juba/rwos R package: R interface to Web of Science Web Services API. Available from <https://github.com/juba/rwos> (accessed September 2017).

Bates D, Mächler M, Bolker B, Walker S. 2014. Fitting linear mixed-effects models using lme4. Preprint. arXiv:1406.5823.

Boakes EH, McGowan PJK, Fuller RA, Chang-qing D, Clark NE, O'Connor K, Mace GM. 2010. Distorted views of biodiversity: spatial and temporal bias in species occurrence data. PLoS Biology **8** DOI: [10.1371/journal.pbio.1000385](https://doi.org/10.1371/journal.pbio.1000385).

Bouché P, Renaud P-C, Lejeune P, Vermeulen C, Froment J-M, Bangara A, Fiongai O, Abdoulaye A, Abakar R, Fay M. 2010. Has the final countdown to wildlife extinction in northern Central African Republic begun? Wildlife extinction Central African Republic. African Journal of Ecology **48**:994–1003.

Brodie JF. 2009. Is research effort allocated efficiently for conservation? Felidae as a global case study. Biodiversity and Conservation **18**:2927–2939.

Brooke ZM, Bielby J, Nambiar K, Carbone C. 2014. Correlates of Research Effort in Carnivores: Body Size, Range Size and Diet Matter. PLoS ONE **9** DOI: [10.1371/journal.pone.0093195](https://doi.org/10.1371/journal.pone.0093195).

Brugière D, Chardonnet B, Scholte P. 2015. Large-Scale Extinction of large carnivores (Lion *Panthera Leo* , Cheetah *Acinonyx Jubatus* and Wild Dog *Lycaon Pictus*) in protected areas of West and Central Africa. Tropical Conservation Science **8**:513–527.

Canby, P. 2016. Shootout in Garamba. The New Yorker, 4 May. Available from <https://www.newyorker.com/news/news-desk/shootout-in-garamba> (accessed September 2017).

Collier P. 2008. The bottom billion: why the poorest countries are failing and what can be done about

it. Oxford University Press, New York.

- Conteh A, Gavin MC, McCarter J. 2017. Assessing the impacts of war on perceived conservation capacity and threats to biodiversity. *Biodiversity and Conservation* **26**:983–996.
- Daskin JH, Pringle RM. 2018. Warfare and wildlife declines in Africa’s protected areas. *Nature* **553**:328–332.
- De Ornellas P, Milner-Gulland EJ, Nicholson E. 2011. The impact of data realities on conservation planning. *Biological Conservation* **144**:1980–1988.
- Douglas-Hamilton I. 1987. African elephants: population trends and their causes. *Oryx* **21**:11.
- Engemann K, Enquist BJ, Sandel B, Boyle B, Jørgensen PM, Morueta-Holme N, Peet RK, Violle C, Svenning J-C. 2015. Limited sampling hampers “big data” estimation of species richness in a tropical biodiversity hotspot. *Ecology and Evolution* **5**:807–820.
- Fay JM. 2007. Ivory wars: last stand in Zakouma. *National Geographic* **211**:34–65.
- Fazey I, Fischer J, Lindenmayer DB. 2005. What do conservation biologists publish? *Biological Conservation* **124**:63–73.
- Fick SE, Hijmans RJ. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* **37**:4302–4315.
- Fisher R, Radford BT, Knowlton N, Brainard RE, Michaelis FB, Caley MJ. 2011. Global mismatch between research effort and conservation needs of tropical coral reefs. *Conservation Letters* **4**:64–72.
- Fox, J. 2003. Effect displays in R for generalised linear models. *Journal of Statistical Software* **8**(15): 1-27.

GADM. 2017. GADM database of global administrative areas. Available from <http://www.gadm.org/> (accessed September 2017).

Gaynor KM, Fiorella KJ, Gregory GH, Kurz DJ, Seto KL, Withey LS, Brashares JS. 2016. War and wildlife: linking armed conflict to conservation. *Frontiers in Ecology and the Environment* **14**:533–542.

Gaston KJ. 2000. Global patterns in biodiversity. *Nature* **405**:220–227.

Giraud, T. 2017. osrm: interface between R and the OpenStreetMap-based routing service OSRM. Available from <https://cran.r-project.org/web/packages/osrm/index.html> (accessed September 2017).

Grenyer R et al. 2006. Global distribution and conservation of rare and threatened vertebrates. *Nature* **444**:93–96.

Hammill E, Tulloch AIT, Possingham HP, Strange N, Wilson KA. 2016. Factoring attitudes towards armed conflict risk into selection of protected areas for conservation. *Nature Communications* **7**:11042.

Hickisch R, Aebischer T. 2013. An update on African elephant *Loxodonta africana* in the Chinko/Mbari Drainage Basin, Central African Republic. Zenodo. Available from <https://zenodo.org/record/997977> (accessed February 2019).

Ibisch PL, Hoffmann MT, Kreft S, Pe'er G, Kati V, Biber-Freudenberger L, DellaSala DA, Vale MM, Hobson PR, Selva N. 2016. A global map of roadless areas and their conservation status. *Science* **354**:1423–1427.

IUCN (International Union for Conservation of Nature) 2017. Red list spatial data. Terrestrial mammals. IUCN, Gland, Switzerland. Available from <https://www.iucnredlist.org/resources/spatial-data-download> (accessed September 2017).

IUCN (International Union for Conservation of Nature), UNEP-WCMC (UN Environment World Conservation Monitoring Centre). 2017. The world database on protected areas. IUCN and UNEP-WCMC, Cambridge, United Kingdom. Available from <https://www.protectedplanet.net/> (accessed September 2017).

Kramer-Schadt S et al. 2013. The importance of correcting for sampling bias in MaxEnt species distribution models. *Diversity and Distributions* **19**:1366–1379.

Lobo JM. 2008. Database records as a surrogate for sampling effort provide higher species richness estimations. *Biodiversity and Conservation* **17**:873–881.

Mannion PD, Upchurch P, Benson RBJ, Goswami A. 2014. The latitudinal biodiversity gradient through deep time. *Trends in Ecology & Evolution* **29**:42–50.

Martin LJ, Blossey B, Ellis E. 2012. Mapping where ecologists work: biases in the global distribution of terrestrial ecological observations. *Frontiers in Ecology and the Environment* **10**:195–201.

Meijaard E, Cardillo M, Meijaard EM, Possingham HP. 2015. Geographic bias in citation rates of conservation research. *Conservation Biology* **29**:920–925.

Meyer C, Jetz W, Guralnick RP, Fritz SA, Kreft H. 2016. Range geometry and socio-economics dominate species-level biases in occurrence information: species-level bias in occurrence records. *Global Ecology and Biogeography* **25**:1181–1193.

NASA (National Aeronautics and Space Administration). 2017a. Version 4 DMSP-OLS nighttime lights time series. Available from <https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html> (accessed August 2017).

NASA (National Aeronautics and Space Administration). 2017b. NRT VIIRS 375 m active fire product VNP14IMGT. Available from <https://firms.modaps.eosdis.nasa.gov/download/> (accessed August 2017).

Nelson, A. (2008). Travel time to major cities: a global map of accessibility. Available from <http://forobs.jrc.ec.europa.eu/products/gam/> (accessed September 2017).

Ondoua OG, Beodo ME, Mambo MJC, Jiagho R, Usongo L, Williamson EA. 2017. An assessment of poaching and wildlife trafficking in the Garamba-Bili-Chinko transboundary landscape. Report TRAFFIC, Cambridge, United Kingdom. Available from <http://static1.1.sqspcdn.com/static/f/157301/27772997/1512751562627/Garamba-Bili-Chinko-summary-EN.pdf?token=CP1dnrHPKFFmlR4LqTpSQYU%2F0Oo%3D> (accessed December 2017)

Reddy S, Dávalos LM. 2003. Geographical sampling bias and its implications for conservation priorities in Africa: sampling bias and conservation in Africa. *Journal of Biogeography* **30**:1719–1727.

Reuters 2017. Web of Science. Available from <http://apps.webofknowledge.com/> (accessed September 2017)

Rondinini C, Wilson KA, Boitani L, Grantham H, Possingham HP. 2006. Tradeoffs of different types of species occurrence data for use in systematic conservation planning: Species data for conservation planning. *Ecology Letters* **9**:1136–1145.

Sastre P, Lobo JM. 2009. Taxonomist survey biases and the unveiling of biodiversity patterns. *Biological Conservation* **142**:462–467.

Tang Y, Horikoshi M, Li W. 2016. ggfortify: unified interface to visualize statistical results of popular

R packages. The R Journal **8**:474.

Economist 2017. The Economist, 9 November. African herders have been pushed into destitution and crime. Available from <https://www.economist.com/news/middle-east-and-africa/21731191-owning-cattle-excellent-way-hide-ill-gotten-wealth-too-african-herders> (accessed November 2017)

Tydecks L, Bremerich V, Jentschke I, Likens GE, Tockner K. 2016. Biological field stations: a global infrastructure for research, education, and public engagement. *BioScience* **66**:164–171.

Velasco D, García-Llorente M, Alonso B, Dolera A, Palomo I, Iniesta-Arandia I, Martín-López B. 2015. Biodiversity conservation research challenges in the 21st century: a review of publishing trends in 2000 and 2011. *Environmental Science & Policy* **54**:90–96.

Venter O, Sanderson EW, Magrach A, Allan JR, Behr J, Jones KR, Possingham HP, Laurance WF, Wood P, Fekete BM. 2016. Sixteen years of change in the global terrestrial human footprint and implications for biodiversity conservation. *Nature Communications* **7**:12558.

Wilson KA, Auerbach NA, Sam K, Magini AG, Moss ASL, Langhans SD, Budiharta S, Terzano D, Meijaard E. 2016. Conservation research is not happening where it is most needed. *PLOS Biology* **14** (e1002413) DOI: .

World Bank. 2017. World Bank open data. Available from <http://data.worldbank.org/> (accessed August 2017).

WorldClim. 2017. WorldClim version 2. Available from <http://worldclim.org/version2> (accessed August 2017).

Yang W, Ma K, Kreft H. 2013. Geographical sampling bias in a large distributional database and its

effects on species richness-environment models. *Journal of Biogeography* **40**:1415–1426.

Table 1: Overview of generated indices used in the analysis of sampling bias in global biodiversity research, their source data, temporal and spatial resolution, and processing heuristic.

Index	Reason	Source data	Type	Temporal resolution	Spatial resolution	Calculation	Heuristic
Imputed number of biodiversity publications	indicator of amount of biodiversity	Web of Science Lite Application Programming Interface	publication query service	1993 - 2016	county	log-transformed sum	Query Web of Science Science Citation Index Expanded and Social Sciences Citation Index for country and province names in various languages. If a country has a total of N publications and C conservation publications, and x publications for a province, the imputed number of biodiversity publications is $INBP=(C/N)*x$
		GADM (Database of Global Administrative Areas) v.2.1.	vector shape	2017	vector		

		R package countrycode	text data	2017	vector		
World road area	indicator of permanent human presence	Ibisch et al. (2016)	vector shape	2016	vector	log- transformed sum	Ibisch et al. (2016) calculated roadless areas. We inverted those to obtain a road index by county.
		GADM (Database of Global Administrative Areas) v.2.1.	vector shape	2017	vector		
Fire cluster	indicator of temporary human presence	NASA (National Aeronautics and Space Administration) VIIRS (Visible Infrared Imaging Radiometer Suite)	vector (centroid of 750m resolution)	jul 2016- jun 17 (1 year)	vector	log- transformed sum	Cluster all fire points in a province by cutting weekly dendrogram of geolocations of these fires at the first level and subsample randomly if a county has more than 20,000 fire points a week.
Night light	indicator of permanent, electrified human presence	NASA (National Aeronautics and Space Administration) night light	raster	2013	NASA night light	mean omitting NAs (where no value is available)	mean by province
Remoteness	Indicator of remoteness and accessibility	Nelson (2008)	Raster	2008	30 arc seconds	log- transformed mean (omitting NAs)	mean of travel time from the province polygon center point to major cities, capitals, airports, and seaports, and the Nelson (2008) data set province average
		Mapbox Directions Service	Travel Time Service	2017	county		
		capital cities from	vector point	2017	vector		

		openstreetmap.org					
		airports from ourairports.com	vector point	2017	vector		
		seaports from the National Geospatial Intelligence Agency	vector point	2017	vector		
Life expectancy	long-term indicator of human development	World Bank Life expectancy at birth, total (years) (SP.DYN.LE00.IN)	text data	1960 - 2010	country	mean (omitting NAs)	mean over 60 years, window that includes values for most countries

Table 2: Model summary for predictors of log-transformed imputed number of publications in the Web of Science (biodiversity and conservation category) as response (3428 observations).

	Predictor	Jurisdiction	Estimate	SE	<i>p</i>
	Number protected areas	province	7.53E-02	1.31E-02	8.65E-09
	Number terrestrial mammals	province	8.14E-02	2.21E-02	2.37E-04
	Protected area, area, cat, iabii	province	7.42E-02	1.16E-02	1.91E-10
	Night light	province	2.10E-01	1.48E-02	2.00E-16
	Life expectancy*	country	1.21E-01	4.94E-02	1.53E-02

	Log road area	province	2.79E-01	2.34E-02	2.00E-16
--	---------------	----------	----------	----------	----------

*Life expectancy is available only at the country level ($n=204$); hence, country was used as grouping variable.

Fig. 1: Geographic density of publications on biodiversity and conservation that mention country and province name (1993-2016) adjusted for size of provinces (square kilometers). An interactive version of this map is available from http://bit.ly/publication_density_map.

Fig. 2: Main effects for mixed model predictors of imputed numbers of biodiversity publications based on model details indicated in Table 2.

Fig. 3: Results of principal components analysis based on the attributes (arrows) of each province (circles). Shading illustrates log-transformed number of publications on biodiversity and conservation imputed (see text for description of relative positions of provinces in the quadrants).

Fig. 4: Location of 100 provinces in quadrant D in Fig. 3 where a bias in number of biodiversity and conservation publications is most apparent. High and low are rankings of countries by their combined principle component score (PC1 positive; PC2 negative). The publications index is the log-transformed imputed number biodiversity and conservation publications.