Social media records hold valuable information for conservation planning

Authors and Affiliations

Shawan Chowdhury*[1,2,3,4], Richard A Fuller[4], Sultan Ahmed[5], Shofiul Alam[5], Corey T Callaghan[6], Priyanka Das[5], Ricardo Correia[7,8,9], Moreno Di Marco[10], Enrico Di Minin[7,8,11], Ivan Jarić[12,13], Mahzabin Muzahid Labi[5], Richard J. Ladle[14,15], Md. Rokonuzzaman[5], Uri Roll[16], Valerio Sbragaglia[17], Asma Siddika[5], Aletta Bonn[1,2,3]

* = corresponding author; contact: dr.shawanchowdhury@gmail.com

[1]Institute of Biodiversity, Friedrich Schiller University Jena, Dornburger Straße 159, 07743 Jena, Germany

[2]Helmholtz Centre for Environmental Research - UFZ, Department of Ecosystem Services, Permoserstr. 15, 04318 Leipzig, Germany

[3]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Puschstr. 4, 04103 Leipzig, Germany

[4]School of Biological Sciences, The University of Queensland, 4072, Australia

[5]Department of Zoology, University of Dhaka, Dhaka 1000, Bangladesh

[6]Department of Wildlife Ecology and Conservation, Fort Lauderdale Research and Education Center, University of Florida, Davie, FL 33314-7719

[7]Department of Geosciences and Geography, University of Helsinki, FI-00014 Helsinki, Finland

[8]Helsinki Institute of Sustainability Science, University of Helsinki, FI-00014 Helsinki, Finland

[9]Biodiversity Unit, University of Turku, 20014 Turku, Finland

[10]Dept of Biology and Biotechnologies, Sapienza University of Rome, viale dell'Università 32, I-00185 Rome, Italy

[11]School of Life Sciences, University of KwaZulu-Natal, Durban 4041, South Africa

[12]Université Paris-Saclay, CNRS, AgroParisTech, Ecologie Systématique Evolution, Orsay, France

[13]Biology Centre of the Czech Academy of Sciences, Institute of Hydrobiology, České Budějovice, Czech Republic

[14]CIBIO/InBIO, Centro de Investigação Em Biodiversidade E Recursos Genéticos, Universidade Do Porto, Campus Agrário de Vairão, 4485-661, Vairão, Portugal

[15]Institute of Biological and Health Sciences, Federal University of Alagoas, Maceió, Alagoas, Brazil

[16]Mitrani Department of Desert Ecology, The Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Midreshet Ben-Gurion 8499000, Israel

[17]Department of Marine Renewable Resources, Institute of Marine Sciences (ICM-CSIC), Barcelona, Spain

Article impact statement

Integrating biodiversity datasets from social media sources could substantially improve our understanding of the natural world.

Keywords

Bangladesh, citizen science, conservation planning, crowdsourcing, iEcology, protected area, social media, tropics, megadiverse countries, Wallacean shortfall

Word count

2943

Running head

Social media – Conservation planning

Acknowledgements

S.C. and A.B. gratefully acknowledge the support of the German Centre for Integrative Biodiversity Research (iDiv) and the sMon project funded by the German Research Foundation (DFG-FZT 118, 202548816).

Code availability

All the R scripts are available in the following public GitHub repository: https://github.com/ShawanChowdhury/SocialMedia_ConservationPlanning.

Social media records hold valuable information for conservation planning

Abstract

Citizen science plays a crucial role in helping monitor biodiversity and inform conservation. With the widespread use of smartphones, many people share biodiversity information on social media, but this information is still not widely used in conservation. Here, focussing on Bangladesh - a tropical mega-diverse and mega-populated country, we examine the potential importance of social media records in conservation decision-making. We show that adding Facebook data to the Global Biodiversity Information Facility (GBIF) data improved the accuracy of conservation planning assessments by identifying additional important conservation areas in the northwest, southeast and centre parts of Bangladesh, extending priority conservation areas by 2,000-5,000 km^2. Community efforts are needed to drive the implementation of the ambitious Kunming-Montreal Global Biodiversity Framework targets, especially in mega-diverse tropical countries with a lack of reliable and up-to-date species distribution data. We highlight that conservation planning can be enhanced by including available data gathered from social media platforms.

Introduction

Earth's biodiversity is unevenly distributed (Pimm et al., 2014). Despite occupying < 2% of the Earth's surface, the tropics contain about 50% of global biodiversity, much of which resides in humid forests (Collen et al., 2008). Most tropical countries have high human population densities, substantial socioeconomic disadvantages, and high dependence on forests (Lewis et al., 2015; Newton et al., 2020). In many tropical regions, forests are over-exploited or are rapidly being converted to agricultural and urban land uses (Bradshaw et al., 2009; Symes et al., 2018; Chowdhury et al., 2021a, b). These multi-faceted human pressures pose an ongoing existential risk to tropical biodiversity (Malhi et al., 2014).

Protected areas (PAs) are the main tool to safeguard biodiversity from these human pressures. PAs play crucial roles in protecting species and populations from extinction (Maxwell et al., 2020; Chowdhury et al., 2022a), and their management can include sustainable land use. The Kunming-Montreal Global Biodiversity Framework (CBD, 2022) includes an ambitious target of expanding the coverage of PAs and other effective area-based conservation measures (OECMs) to 30% by 2030, emphasising area-based conservation approaches as a key means to maintain species and ecosystem functions. The effectiveness of such an approach largely depends on maximising biodiversity protection in PAs, requiring detailed records of the distribution of species. While such data are often available for Europe and North America, tropical taxa are typically less well-sampled (Di Marco et al., 2017; Troudet et al., 2017).

Citizen science is playing a crucial role in filling global biodiversity knowledge gaps (Di Minin et al., 2015; Chandler et al., 2017; Callaghan et al., 2021, 2022), and even in Europe, around 80%–90% of biodiversity observational records are collected by dedicated volunteers (Schmeller et al., 2009). Amateur (and professional) naturalists are increasingly taking advantage of expanded internet coverage and the photographic capacity of mobile devices to share their observations online (Andrachuk et al., 2019; Marcenò et al., 2021). Consequently, the amount of biodiversity data from citizen science in the Global Biodiversity Information Facility (GBIF) is sharply increasing, although with a bias towards Europe and North America (Hughes et al., 2021). Due to the increasing popularity of social media (e.g., Facebook, Flickr), millions of people post biodiversity photographs (Toivonen et al., 2019). If these biodiversity observation records can also be captured and mobilised, this could enhance existing knowledge of tropical species distributions with the potential to vastly improve conservation assessments (Toivonen et al., 2019; Jarić et al., 2020). Conservation science has so far utilised social media data only in some instances, such as mapping ecosystem services, promoting conservation through marketing and education, monitoring species, ecosystems, and management, and facilitating conservation communication (Di Minin et al., 2015). Here, using a tropical South Asian mega-populated country, Bangladesh, we test whether social media data can directly contribute to conservation decision-making.

Bangladesh is part of the Indo-Burma and Indo-Malayan biodiversity hotspots (Chowdhury et al., 2021a) and is home to many globally charismatic species, including Royal Bengal Tiger (*Panthera tigris tigris*), spoon-billed sandpiper (*Calidris pygmaea*) and the Ganges River dolphin (*Platanista gangetica*). About 25% of assessed species in Bangladesh are threatened with extinction (IUCN Bangladesh, 2015). Biodiversity data from Bangladesh are scarce in GBIF (0.0001%), like many other tropical countries; however, there is an active community of amateur photographers whose images, posted on social media platforms such as Facebook, often contain biodiversity information. A recent study captured 7,096 records of butterflies from Bangladesh posted on Facebook (compared to only 205 observations on GBIF; Chowdhury et al. 2021b).

Here, we examine the importance of using social media records to inform conservation decision-making, using Bangladesh as a case study. To achieve this, we collated species distribution records for the most photographed group of vertebrates (birds) and invertebrates (butterflies) from Facebook and GBIF. We fitted species distribution models and evaluated how existing PAs cover the predicted species' geographic range. We further calculated species representation targets (within a formal conservation prioritisation scheme) to identify priority areas for future PA establishment and conservation actions in Bangladesh. Finally, we investigated differences in selecting essential conservation areas between GBIF-only and combined data (GBIF and Facebook). We reveal how social media data could complement and expand existing biodiversity data and, consequently, conservation planning.

Methods

Data

We compiled a comprehensive checklist of birds and butterflies of Bangladesh from the most recent national Red List data book (Total = 871 species; Birds = 566 species, Butterflies = 305 species; IUCN Bangladesh, 2015). We collected climatic data from the WordClim database (http://www.worldclim.com/version2) at the finest resolution (0.693 km^2; 833 m * 833 m), the distribution of the current PAs in Bangladesh (UNEP-WCMC, 2021) using the 'wdpar' R package, and land-use data from Copernicus Global Land Service (Buchhorn et al., 2020). For the spatial data, we followed two steps: first, we collected species distribution records from the GBIF (https://www.gbif.org/) and then collated and georeferenced biodiversity data from Facebook.

Initially, we downloaded spatial distribution records for the birds and butterflies of Bangladesh from the GBIF using the rgbif package (Chamberlain et al., 2022) in R (R version 4.0.4; R Core Team, 2021). Being the largest global biodiversity data infrastructure network, GBIF compiles occurrence records from various sources - from museum specimens to citizen science records. To avoid repetition, we did not collect data from other biodiversity repositories that feed data into GBIF (e.g., iNaturalist).

Finally, we collected species distribution records from Facebook from our previous work (Chowdhury et al., 2022). These records were obtained by searching for species distribution records in two Facebook groups: Birds Bangladesh (https://www.facebook.com/groups/2403154788) and Butterfly Bangladesh (https://www.facebook.com/groups/488719627817749). In each group, we explored data by species common name obtained from IUCN Bangladesh (2015), double-checked the identification in each photograph, and extracted the species details (taxonomic information, location, date, and photographers). Afterwards, for each observation, we georeferenced the location using Google Maps (https://maps.google.com/). We excluded pictures if the identification was incomplete (not up to species level) or wrong, if the photograph did not allow clear taxonomic identification, if the location was unspecified, or could not be accurately determined.

While different social media channels (e.g., Facebook, Flickr, WhatsApp) can be efficient sources of obtaining biodiversity data (Toivonen et al., 2019), we only considered Facebook for our study. This was because i) Facebook is the most popular social media channel for the photographers of Bangladesh, and ii) the locality information is typically much more vague in other social media channels (e.g., Twitter). When sharing biodiversity photographs in Facebook groups, photographers are required by group rules to specify the location so that group members can evaluate the records (Chowdhury et al., 2021a).

Data cleaning

We cleaned GBIF data using the CoordinateCleaner R package (Zizka et al., 2019). We removed duplicate records, precision uncertainty over 10 km, imprecise coordinates (zero coordinates, integers, records in oceans), and invalid coordinates (where the specified locality was incompatible with the coordinates given.

To address sampling bias, we followed two steps. First, we spatially thinned the combined data using the spThin R package (Aiello-Lammens et al., 2015). We only considered a single occurrence record at 0.693 km^2 (833 m * 833 m) resolution for each species. We followed the same process for the GBIF dataset.

We cleaned the PA data following a globally accepted method (Butchart et al., 2015). We rasterised the protected boundaries at 0.693 km^2 (833 m * 833 m) resolution using the 'fasterize' R package (Ross, 2020).

We checked collinearity among the WorldClim variables and removed highly correlated ($r > 0.75$) variables. This way, we removed 11 of the 19 climatic variables.

Cleaning protected area

We cleaned the PA data following a globally accepted method (Butchart et al., 2015). Namely, we reprojected the data into an equal-area coordinate system (World Behrmann; ESRI: 54017), excluded

UNESCO biosphere reserves and sites with unknown or proposed status, created buffers around PAs denoted as point localities, and expanded them to their reported extent. The cleaned PA dataset resulted in boundaries for 42 PAs.

Habitat suitability maps

We fitted MaxEnt species distribution models to generate habitat suitability maps using the ENMEval package in R (Muscarella et al., 2014). We ran the model separately for GBIF-only data and combined data using the following method.

We fitted species distribution models for each species using 9 predictor variables (8 climatic and 1 land-use) with 10-fold cross-validation and 5,000 randomly generated background records at 0.693 km^2 resolution. As Bangladesh is a small country, we used 5,000 instead of the typical 10,000 background records. We generated folds by overlaying the presence and background records with a spatial grid to control sampling bias and spatial auto-correlation on model performance. We assigned the records to grid cells and then randomly assigned grid cells to particular folds (Muscarella et al., 2014). To further improve model performance, we performed a calibration procedure by fitting the model under different combinations of parameters. Specifically, we fitted the model under six feature class combinations ("L", "LQ", "H", "LQH", "LQHP", and "LQHPT") and eight different regularisation multipliers (0.5-4, with 0.5 intervals). We evaluated the models using the AUC (Area Under a Receiver Operating Characteristic –ROC–z Curve). After identifying the best model for each species, we used them to generate continuous habitat suitability maps across the study area. We then applied thresholds to convert the continuous values into binary values, resulting in maps that denote the presence or absence of suitable habitat conditions. The threshold values were specified based on maximising the sum of the sensitivity and specificity statistics (Liu et al., 2016). Since the best models all had an AUC statistic greater than 0.7 (mean AUC = 0.92), we are confident that they are suitable to address the aims of our study. We also checked the suitability distribution of the predicted maps, and the prediction was good with low omission and commission errors.

We used the binary habitat suitability maps for subsequent analysis. We had 470 species (GBIF data) and 698 species (combined data) for the final analysis. We extracted built-up areas from the land-use map and removed suitable habitats within these areas for each species for both Facebook and combined datasets.

Protected area coverage

To evaluate the extent to which existing PAs in Bangladesh overlapped with biodiversity, we overlaid the species' binary habitat suitability maps with the PA data and measured the percentage of suitable habitat within the existing PAs. Afterwards, we compared the percent level of coverage to a target threshold (termed "representation target"). We set the targets following a modified version of standard practices for global analysis (Butchart et al., 2015). We set the target at 100% for species with a distribution of <1,000 km^2, 10% for those with 148,460 km^2 (country area) and for the intermittent values, we log-linearly interpolated the targets using the 'prioritizr' R package (Hanson et al., 2022).

Spatial conservation prioritisation

We identified priority areas that most efficiently fill shortfalls in the existing PA system using (i) the GBIF data, and (ii) the combined data. For this, we generated a prioritisation based on the minimum set formulation of the reserve selection problem, where the grid cells were used as planning units. We generated these prioritizations with the species' binary habitat suitability maps and the representation targets we used in the previous step to assess the performance of existing PAs. To account for opportunity costs associated with implementing conservation areas, we considered the human footprint index (Venter et al., 2018) as cost data (Butchart et al., 2015). Additionally, to ensure that priority areas complement existing PAs, existing PAs were locked in. These analyses were completed using the 'prioritizr' R package (Hanson et al., 2022) and Gurobi (version 8.1.0; Gurobi Optimization, LLC, 2021), and by solving optimisation routines to optimality. After generating the prioritisation, we overlaid it with land-use data to facilitate interpretation.

To identify the most important priority areas in the analysis, we ran an irreplaceability analysis using the rarity-weighted richness metric for each planning unit selected in a solution using the 'prioritizr' R package. While running the irreplaceability analysis, to quantify the importance of planning units, we used the Ferrier score (Ferrier et al., 2000).

Results

Data distribution

Our cleaned combined dataset included 47,077 georeferenced records for 472 species of birds (41,476 records) and 226 species of butterflies (5,601 records). We obtained 49% of the records from GBIF (N = 22,885), including 540 species (428 birds and 112 butterflies), and 51% of the records from Facebook, including 158 new species (44 birds and 114 butterflies). Facebook data provided substantial variations across species and taxa (Figure 1a; supplementary Figure S1). For butterflies, the average number of occurrence records (per species) jumped from 2 to 25 after including Facebook data, while GBIF data never represented more than 21% of species' records (Figure 1a). For birds, the inclusion of Facebook records raised the average number of occurrence records per species from 48 to 88. While there were no butterfly species with GBIF-only distribution data, there were 18 bird species for which we obtained data only from GBIF (Figure 4a).

Habitat suitability maps

Overall model performance was good using either dataset. After including Facebook data, the average AUC score only increased from 0.92 to 0.93. However, using GBIF-only data led to 228 species (33%) not being included in the modelling due to having a limited number of spatial observation records; this was especially true for butterflies (161 species, 71% of butterflies) and less so for birds (67 species, 14% of birds) (Figure 1b).

Spatial conservation prioritisation

Using GBIF-only data, the spatial prioritisation process identified that 25.33 % of the country's area was required for birds (37,605 km^2) and 19.97 % for butterflies (29,647 km^2) to meet the target conservation coverage (see Methods). After adding Facebook data, the prioritisation process identified an area of 26.75 % for birds (increased by 2,108 km^2) and 23.39 % for butterflies (increased by 5,077 km^2). For birds, by using GBIF-only data, the prioritisation process missed many important areas distributed in the north, northeast and southeast parts of Bangladesh compared to the combined data. However, when considering the current land-use patterns across Bangladesh (Figure 2a), there were no substantial differences in the proportion of land-cover type selection.

For butterflies, priority areas identified using GBIF-only data also missed many parts from the northwest, southeast and centre parts of Bangladesh. However, similar to birds, there were little differences in land cover types between the two schemes. The proportion of priority areas peaked along croplands and forests and was lowest for shrublands and herbaceous vegetation (Figure 2b,c supplementary figure S3). In both GBIF-only and combined data, the proportion of priority areas was highest for croplands and forests and lowest for shrublands and herbaceous vegetation (Figure 2a,b, supplementary Figure S3).

To determine if the prioritisation process identified more areas simply due to the inclusion of more species, we reran the prioritisation by only including species for which we obtained suitability maps using both approaches. We found that the number of important conservation areas identified by the prioritisation process was lower – compared to the combined data with all species – for birds and butterflies (see the supplementary Figure S2).

Given our definition of a cost surface based on HFP, priority areas were primarily distributed in places with a low level of anthropogenic impact (see Methods) using both GBIF-only and combined data for both birds and butterflies. However, with the inclusion of Facebook data, the priority areas' mean and median HFP index increased slightly (supplementary Figure S4). After adding Facebook data, butterflies' mean and median HFP index of the priority areas increased from 12.8 to 13.8 and 11.6 to 14.3. The mean and median HFP index for birds increased similarly from 13.7 to 14 and from 14.3 to 15.8.

Irreplaceability score

To identify the most important priority areas in the prioritisation process, we ran an irreplaceability analysis. While most priority areas had relatively low irreplaceability, the scores improved markedly after adding Facebook data. For birds, 25 % of priority areas had a score > 0.0001 using combined data, compared to only 0.00009% with GBIF-only data (Figure 3a, b). For butterflies, 25 % of priority areas had an irreplaceability score > 0.0004 using combined data, compared to 0.0001% with the GBIF-only data (Figure 3c, d).

While, for birds (Figure 3a, b), additional Facebook data caused little difference in identifying the most crucial priority areas (top 10%), we obtained marked differences for butterflies (Figure 3c, d). With the addition of Facebook records, for birds, irreplaceability scores increased in the north and southwest parts of Bangladesh, whereas centre, north, and northwest regions had the highest scores in both scenarios (Figure 3a, b). For butterflies, the irreplaceability scores increased substantially in the centre, northeast, and east parts of Bangladesh (Figure 3c, d).

Discussion

Bangladesh, like many tropical countries, is highly biodiverse. Yet knowledge of most of its species' distribution is limited (IUCN Bangladesh, 2015). The ubiquitous availability of digital phones and cameras creates abundant opportunities for people in less-represented countries to post their biodiversity photographs on social media. Here we compared biodiversity data between GBIF and a combined GBIF/Facebook dataset. This demonstrated that data obtained from social media have a significant capacity to inform important conservation decision-making (priority areas identified increased by 2,000 - 5,000 km^2). A prioritisation scheme, that included the Facebook data, identified more areas, especially due to the inclusion of more species in the analysis. The number and location of identified conservation priority areas increased sharply after adding Facebook data, and there were marked differences in the most valuable irreplaceable areas (between schemes with and without these data), especially for butterflies.

The increasing popularity of citizen science has greatly improved our understanding of species distributions in recent years. There has been a 12-fold increase in biodiversity data in GBIF since 2007 (Heberling et al., 2021), albeit mostly in the Global North. In countries with a lack of natural-history museums and systematic monitoring schemes, like Bangladesh, citizen science can play an especially efficient role in recording biodiversity data. When we included Facebook data, our occurrence records doubled. Remarkably, data for more than two-thirds of butterfly species were only available from Facebook. During our initial data collation for birds in GBIF, we did not collect their original data source; however, based on a random check, the majority came from eBird, and the contribution of museum data was negligible. Bangladesh has many active eBird users, but there are not many butterfly enthusiasts that use specialised butterfly citizen science applications, which probably caused the difference in GBIF records between birds and butterflies (41476 vs. 5601, respectively).

By including observation records from Facebook, our systematic conservation planning approaches identified many new important areas from the northeastern and southeastern parts of Bangladesh. Despite being home to many charismatic species and biodiversity hotspots, the importance of these areas for biodiversity conservation remains unnoticed. These areas are occupied mainly by indigenous communities, are distant from metropolitan areas, and most people there have not been habituated to using citizen science applications with dedicated biodiversity monitoring schemes. Many people living in these areas are, however, Facebook users. Moreover, wildlife photographers from other parts of Bangladesh often visit these regions and share their photographs on Facebook. Therefore, our results highlight the great utility of combining biodiversity repositories and social media data for conservation monitoring and planning, across scales, especially in less monitored regions (Kelling et al., 2019).

Limitations and caveats

Biodiversity monitoring needs a culture of integration (Kühle et al., 2022), and it will be important to align different data sources. Here using biodiversity data from two sources, we show how additional records from social media can influence conservation decisions; however, this result should be interpreted carefully. For example, the citizen science data is highly biased and largely centred around major cities, which might have an impact on our results. However, we followed a range of approaches to control the survey bias (e.g., spatial thinning) and model prediction (e.g., the 'checkerboard2' evaluation method to control biased sampling).

While social media can play an important role in supporting biodiversity conservation assessments, there remain considerable challenges to capturing and collating such data. First, while biodiversity data can be harvested from different social media channels (Flickr, WhatsApp), we only chose Facebook for our study, in our case because Facebook groups in Bangladesh are regularly monitored by group moderators, unlike other social media. Second, capturing biodiversity data from Facebook is very time-consuming, taking about 380 hours to harvest data for all 680 species in our study (Chowdhury et al., 2022b). Third, Facebook photographs often do not contain specific geolocation information, resulting in frequent coordinate uncertainty when georeferencing. Finally, accurate species identification using Facebook photographs requires high-quality pictures and a high level of taxonomic expertise.

Taking photographs rich in taxonomic information is difficult, and many species remain unidentified as a consequence. To enhance semi-structured monitoring in citizen science (Kelling et al., 2019), Facebook group moderators could help train recorders, and photos could have automated GPS records attached. In addition, citizen science records could be enhanced using novel technologies such as camera traps and using artificial intelligence for automated image recognition (van Klink et al., 2022). Furthermore platforms more narrowly dedicated to recording biodiversity data (e.g., eButterfly, Flora Incognita, iNaturalist), could be used to augment Facebook (and other social media) data. In turn, information from such, more dedicated, sources could be used to develop and train deep-learning image classification and identification models (Jarić et al., 2020), especially for lesser-known tropical species. Overall, promoting the importance of citizen science in biodiversity conservation and the broader availability of digital apps can generate extensive data from remote areas. Moreover, citizen science can also heighten awareness of biodiversity and help engender a sense of social responsibility or social licence for conservation (Kelly et al., 2019).

Conclusions

Our understanding of tropical biodiversity remains limited. Yet, with the increasing popularity of mobile phones and social media platforms, millions of users habitually share valuable biodiversity information through photographs. Such information, if carefully harvested and collated, could significantly decrease the Wallacean shortfall (Hortal et al. 2015). With the addition of biodiversity data collected from Facebook (or similar sources), knowledge of many range-restricted species can be significantly improved and inform more effective conservation management. While our study is focussed on Bangladesh, its methods could be applied to many tropical developing countries with sufficiently good internet penetration and an active culture of social media use. The Kunming-Montreal Global Biodiversity Framework prioritises area-based conservation approaches, placing a premium on rapidly improving our knowledge of species distributions. Combining data from multiple repositories, including social media, should thus be a priority to improve the quality of large-scale conservation planning. In short, if the limitations of capturing, cleaning and collating biodiversity information from social media platforms can be overcome, there is an enormous potential for improving biodiversity conservation, both globally and especially in tropical megadiverse countries.

Data availability

We obtained the species distribution data in two ways: from GBIF and Facebook. Both GBIF data (GBIF, 2022) and Facebook data (Chowdhury et al., 2022b) are publicly available.

References

Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B., & Anderson, R. P. (2015). spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*, *38*, 541-545. https://doi.org/10.1111/ecog.01132.

Andrachuk, M., Marschke, M., Hings, C., & Armitage, D. (2019). Smartphone technologies supporting community-based environmental monitoring and implementation: a systematic scoping review. *Biological Conservation*, *237*, 430-442. https://doi.org/10.1016/j.biocon.2019.07.026.

Bradshaw, C. J., Sodhi, N. S., & Brook, B. W. (2009). Tropical turmoil: a biodiversity tragedy in progress. *Frontiers in Ecology and the Environment*, *7*, 79-87. https://doi.org/10.1890/070193.

Buchhorn, M., Smets, B., Bertels, L., De Roo, B., Lesiv, M., Tsendbazar, N. - E. et al. (2020). Copernicus Global Land Service: Land Cover 100m: collection 3: epoch 2019: Globe 2020. Available at: 10.5281/zenodo.3939050.

Butchart, S. H., Clarke, M., Smith, R. J., Sykes, R. E., Scharlemann, J. P., Harfoot, M., ... & Burgess, N. D. (2015). Shortfalls and solutions for meeting national and global conservation area targets. *Conservation Letters*, *8*, 329-337. https://doi.org/10.1111/conl.12158.

Callaghan, C. T., Mesaglio, T., Ascher, J. S., Brooks, T. M., Cabras, A. A., Chandler, M., ... & Young, A. N. (2022). The benefits of contributing to the citizen science platform iNaturalist as an identifier. *PLoS Biology*, *20*, e3001843. https://doi.org/10.1371/journal.pbio.3001843.

Callaghan, C. T., Poore, A. G., Mesaglio, T., Moles, A. T., Nakagawa, S., Roberts, C., ... & Cornwell, W. K. (2021). Three frontiers for the future of biodiversity research using citizen science data. *BioScience*, *71*, 55-63. https://doi.org/10.1093/biosci/biaa131.

CBD (2022) Kunming-Montreal Global biodiversity framework. Draft decision submitted by the President CBD/COP/15/L.25, 18 December 2022. Convention on Biological Diversity, Montreal, Canada. https://www.cbd.int/conferences/2021-2022/cop-15/documents.

Chamberlain, S., Barve, V., Mcglinn, D., Oldoni, D., Desmet, P., Geffert, L. & Ram, K. (2022). *rgbif: Interface to the Global Biodiversity Information Facility API*. R package version 3.7.2, https://CRAN.R-project.org/package=rgbif.

Chandler, M., See, L., Copas, K., Bonde, A. M., López, B. C., Danielsen, F., ... & Turak, E. (2017). Contribution of citizen science towards international biodiversity monitoring. *Biological Conservation*, *213*, 280-294. https://doi.org/10.1016/j.biocon.2016.09.004.

Chowdhury, S., Aich, U., Rokonuzzaman, M., Alam, S., Das, P., Siddika, A.,.., & Fuller, R. A. (2022). Spatial occurrence data for the animals of Bangladesh derived from Facebook. PANGAEA, https://doi.org/10.1594/PANGAEA.948104.

Chowdhury, S., Alam, S., Chowdhury, S. U., Rokonuzzaman, M., Shahriar, S. A., Shome, A. R., & Fuller, R. A. (2021b). Butterflies are weakly protected in a mega-populated country, Bangladesh. *Global Ecology and Conservation*, *26*, e01484. https://doi.org/10.1016/j.gecco.2021.e01484.

Chowdhury, S., Alam, S., Labi, M. M., Khan, N., Rokonuzzaman, M., Biswas, D., ... & Fuller, R. A. (2021a). Protected areas in South Asia: Status and prospects. *Science of the Total Environment*, 152316. https://doi.org/10.1016/j.scitotenv.2021.152316.

Chowdhury, S., Jennions, M. D., Zalucki, M. P., Maron, M., Watson, J. E., & Fuller, R. A. (2022a). Protected areas and the future of insect conservation. *Trends in Ecology & Evolution*, *38*, 85-95. https://doi.org/10.1016/j.tree.2022.09.004.

Collen, B., Ram, M., Zamin, T., & McRae, L. (2008). The tropical biodiversity data gap: addressing disparity in global monitoring. *Tropical Conservation Science*, *1*(2), 75-88. https://doi.org/10.1177/194008290800100202.

Di Marco, M., Chapman, S., Althor, G., Kearney, S., Besancon, C., Butt, N., ... & Watson, J. E. M. (2017). Changing trends and persisting biases in three decades of conservation science. *Global Ecology and Conservation*, *10*, 32-42. https://doi.org/10.1016/j.gecco.2017.01.008.

Di Minin, Enrico, Henrikki Tenkanen, and Tuuli Toivonen. (2015). Prospects and challenges for social media data in conservation science. *Frontiers in Environmental Science 3*, 63. https://doi.org/10.3389/fenvs.2015.00063.

Ferrier, S., Pressey, R. L., & Barrett, T. W. (2000). A new predictor of the irreplaceability of areas for achieving a conservation goal, its application to real-world planning, and a research agenda for further refinement. *Biological Conservation*, *93*, 303-325. https://doi.org/10.1016/S0006-3207(99)00149-4.

GBIF. (2022). Derived dataset GBIF.org (11 April 2022), Filtered export of GBIF occurrence data. https://doi.org/10.15468/dd.rghepr.

Gurobi Optimization, LLC. (2021). gurobi: Gurobi Optimizer 8.1 interface. R package version 8.1-0. Retrieved from http://www.gurobi.com.

Hanson J. O. (2021). wdpar: Interface to the World Database on Protected Areas. R package version 1.3.2. https://CRAN.R-project.org/package=wdpar.

Hanson JO, Schuster R, Morrell N, Strimas-Mackey M, Edwards BPM, Watts ME, Arcese P, Bennett J, Possingham HP (2022). prioritizr: Systematic Conservation Prioritization in R. R package version 7.2.2. Available at https://CRAN.R-project.org/package=prioritizr.

Heberling, J. M., Miller, J. T., Noesgaard, D., Weingart, S. B., & Schigel, D. (2021). Data integration enables global biodiversity synthesis. *Proceedings of the National Academy of Sciences*, *118*, e2018093118. https://doi.org/10.1073/pnas.2018093118.

Hijmans, R. J., Van Etten, J., Cheng, J., Mattiuzzi, M., Sumner, M., Greenberg, J. A., ...& Hijmans, M. R. J. (2015). Package ‘raster’. *R package*, 734.

Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven shortfalls that beset large-scale knowledge of biodiversity. *Annual Review of Ecology, Evolution, and Systematics, 46*, 523-549. https://doi.org/10.1146/annurev-ecolsys-112414-054400.

Hughes, A. C., Orr, M. C., Ma, K., Costello, M. J., Waller, J., Provoost, P., ... & Qiao, H. (2021). Sampling biases shape our view of the natural world. *Ecography*, *44*, 1259-1269. https://doi.org/10.1111/ecog.05926.

IUCN Bangladesh. (2015). Red list of Bangladesh: a brief on assessment result 2015, IUCN, International Union for Conservation of Nature, Bangladesh Country Office, Dhaka, Bangladesh.

Jarić, I., Correia, R. A., Brook, B. W., Buettel, J. C., Courchamp, F., Di Minin, E., ... & Roll, U. (2020). iEcology: harnessing large online resources to generate ecological insights. *Trends in Ecology & Evolution*, *35*, 630-639. https://doi.org/10.1016/j.tree.2020.03.003.

Kelling, S., Johnston, A., Bonn, A., Fink, D., Ruiz-Gutierrez, V., Bonney, R., ... & Guralnick, R. (2019). Using semistructured surveys to improve citizen science data for monitoring biodiversity. *BioScience*, *69*, 170-179. https://doi.org/10.1093/biosci/biz010.

Kelly, R., Fleming, A., Pecl, G., Richter, A. & Bonn, A. (2019) Social licence through citizen science: A tool for marine conservation. *Ecology & Society*, 14 16. https://doi.org/10.5751/ES-10704-240116

Kühl, H.S., Bowler, D.E., Bösch, L., Bruelheide, H., Dauber, J., Eichenberg, D., Eisenhauer, N., Fernández, N., Guerra, C.A., Henle, K., Herbinger, I., Isaac, N.J.B., Jansen, F., König-Ries, B., Kühn, I., Nilsen, E.B., Pe'er, G., Richter, A., Schulte, R., Settele, J., van Dam, N.M., Voigt, M., Wägele, W.J., Wirth, C. & Bonn, A. (2020) Effective biodiversity monitoring needs a culture of integration. One Earth, 3, 462-474. https://doi.org/10.1016/j.oneear.2020.09.010

Lewis, S. L., Edwards, D. P., & Galbraith, D. (2015). Increasing human dominance of tropical forests. *Science*, *349*, 827-832. https://doi.org/10.1126/science.aaa9932.

Liu, C., Newell, G., & White, M. (2016). On the selection of thresholds for predicting species occurrence with presence-only data. *Ecology and Evolution*, 6, 337-348. https://doi.org/10.1002/ece3.1878.

Malhi, Y., Gardner, T. A., Goldsmith, G. R., Silman, M. R., & Zelazowski, P. (2014). Tropical forests in the Anthropocene. *Annual Review of Environment and Resources*, *39*, 125-159. https://doi.org/10.1146/annurev-environ-030713-155141.

Marcenò, C., Padullés Cubino, J., Chytrý, M., Genduso, E., Salemi, D., La Rosa, A., ... & Guarino, R. (2021). Facebook groups as citizen science tools for plant species monitoring. *Journal of Applied Ecology*, *58*, 2018-2028. https://doi.org/10.1111/1365-2664.13896.

Maxwell, S. L., Cazalis, V., Dudley, N., Hoffmann, M., Rodrigues, A. S., Stolton, S., ... & Watson, J. E. M. (2020). Area-based conservation in the twenty-first century. *Nature*, *586*, 217-227. https://doi.org/10.1038/s41586-020-2773-z.

Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., Uriarte, M., & Anderson, R. P. (2014). ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods in Ecology and Evolution*, 5, 1198-1205. https://doi.org/10.1111/2041-210X.12261.

Newton, P., Kinzer, A. T., Miller, D. C., Oldekop, J. A., & Agrawal, A. (2020). The number and spatial distribution of forest-proximate people globally. *One Earth*, *3*, 363-370. https://doi.org/10.1016/j.oneear.2020.08.016.

Pimm, S. L., Jenkins, C. N., Abell, R., Brooks, T. M., Gittleman, J. L., Joppa, L. N., ... & Sexton, J. O. (2014). The biodiversity of species and their rates of extinction, distribution, and protection. *Science*, *344*, 1246752. https://doi.org/10.1126/science.1246752.

Pocock, M.J., Roy, H.E., August, T., Kuria, A., Barasa, F., Bett, J., Githiru, M., Kairo, J., Kimani, J. & Kinuthia, W. (2019). Developing the global potential of citizen science: Assessing opportunities that benefit people, society and the environment in East Africa. *Journal of Applied Ecology*, *56*, 274-281. https://doi.org/10.1111/1365-2664.13279.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ross, N. (2020). fasterize: Fast polygon to raster conversion. *R package version*, *1*(3).

Schmeller, D.S., Henry, P.Y., Julliard, R., Gruber, B., Clobert, J., Dziock, F., Lengyel, S., Nowicki, P., De´ri, E., and Budrys, E. (2009). Advantages of volunteer-based biodiversity monitoring in Europe. *Conservation Biology*, *23*, 307–316. https://doi.org/10.1111/j.1523-1739.2008.01125.x.

Symes, W. S., Edwards, D. P., Miettinen, J., Rheindt, F. E., & Carrasco, L. R. (2018). Combined impacts of deforestation and wildlife trade on tropical biodiversity are severely underestimated. *Nature Communications*, *9*, 1-9. https://doi.org/10.1038/s41467-018-06579-2.

Toivonen, T., Heikinheimo, V., Fink, C., Hausmann, A., Hiippala, T., Järv, O., ... & Di Minin, E. (2019). Social media data for conservation science: A methodological overview. *Biological Conservation*, *233*, 298-315. https://doi.org/10.1016/j.biocon.2019.01.023.

Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R., & Legendre, F. (2017). Taxonomic bias in biodiversity data and societal preferences. *Scientific Reports*, *7*, 1-14. https://doi.org/10.1038/s41598-017-09084-6.

UNEP-WCMC. (2021). Protected Area Profile for Bangladesh from the World Database of Protected Areas, Cambridge, UK: UNEP-WCMC and IUCN. (2020). Available at: www.protectedplanet.net.

van Klink, R., August, T., Bas, Y., Bodesheim, P., Bonn, A., Fossøy, F., ... & Bowler, D. E. (2022) Emerging technologies revolutionise insect ecology and monitoring. *Trends in Ecology & Evolution*, *37*, 872-885. https://doi.org/10.1016/j.tree.2022.06.001

Venter, O., Sanderson, E. W., Magrach, A., Allan, J. R., Beher, J., Jones, K. R., ... & Watson, J. E. M. (2018). Last of the wild project, version 3 (LWP-3): 2009 human footprint, 2018 release. *Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC)*, *10*, H46T0JQ4.

Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Duarte Ritter, C., Edler, D., ... & Antonelli, A. (2019). CoordinateCleaner: Standardised cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution*, *10*(5), 744-751.

List of Figures

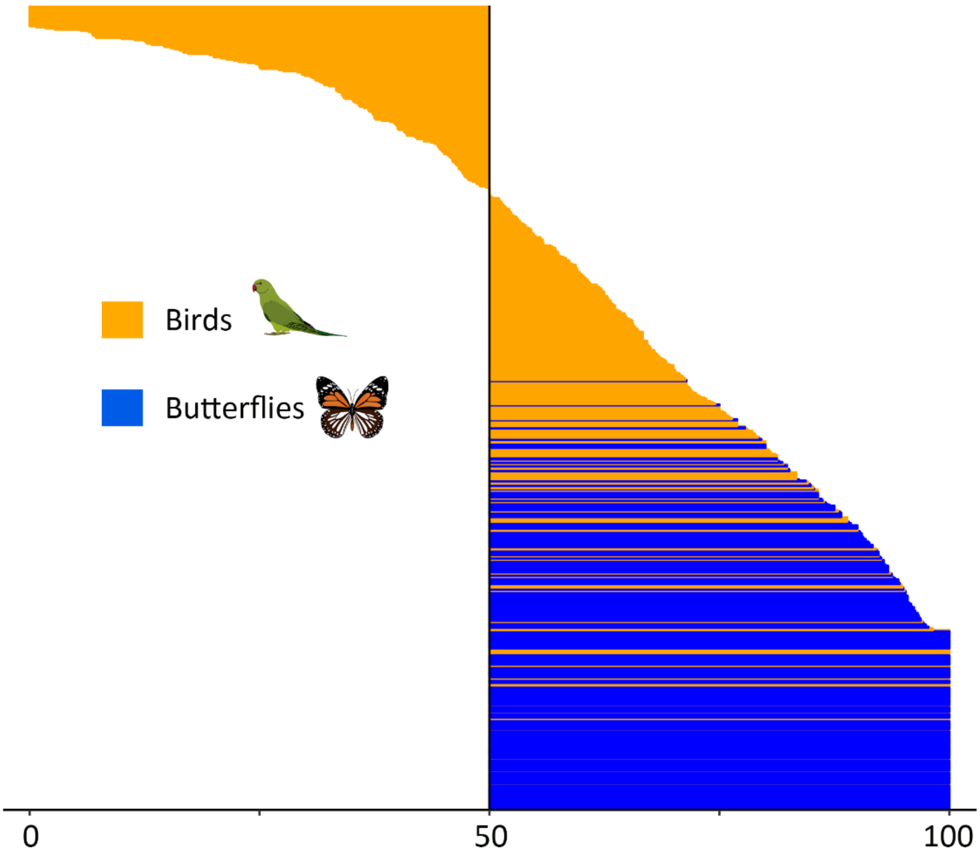
Figure 1. Gaps in restricting biodiversity data to the Global Biodiversity Information Facility (GBIF). Here, (a) shows the percentages of biodiversity records that went missing by excluding Facebook data (e.g., 0 indicates 0% data missing, 100 indicates 100% data missing, and 50 indicates we obtained the same number of records from Facebook and GBIF), and (b) compares the number of species for which we obtained habitat suitability maps using GBIF and combined data (all records).

Figure 2. (a) A land-cover map of Bangladesh; (b) the spatial mismatch in identifying priority cells for birds; and (c) butterflies.

Figure 3. Maps comparing the differences in classifying the importance of priority cells for birds (a & b) and butterflies (b & d) in Bangladesh. Here, red rectangles indicate areas with key differences.

Figure 1

a) Percentages of species occurrence records missing using GBIF-only data



b) The number of species for which we obtained suitability maps in comparison to the total available species

Group	Total species	With GBIF-only data	With combined data
Birds	566	405	472
Butterflies	305	65	226

Figure 2

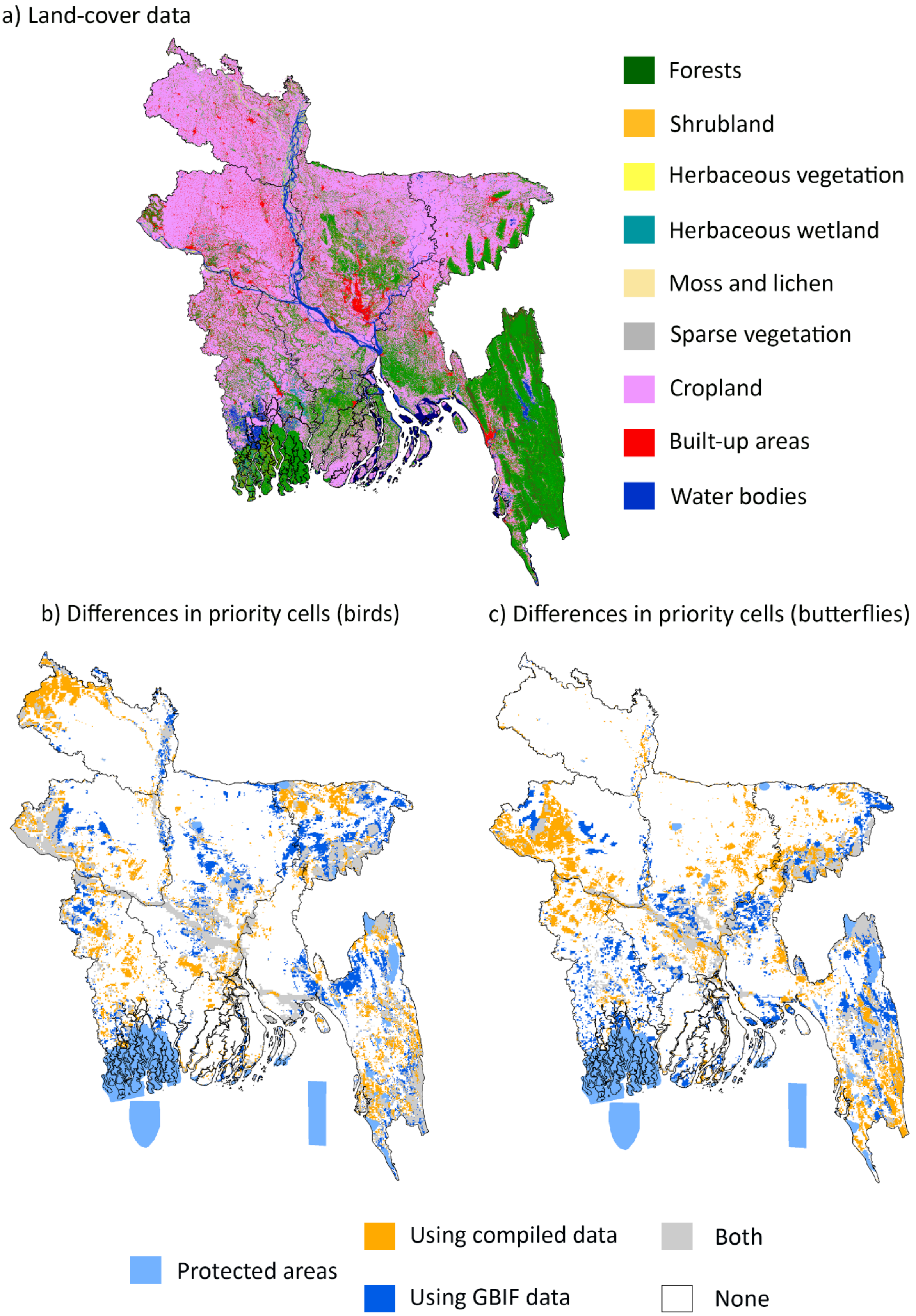
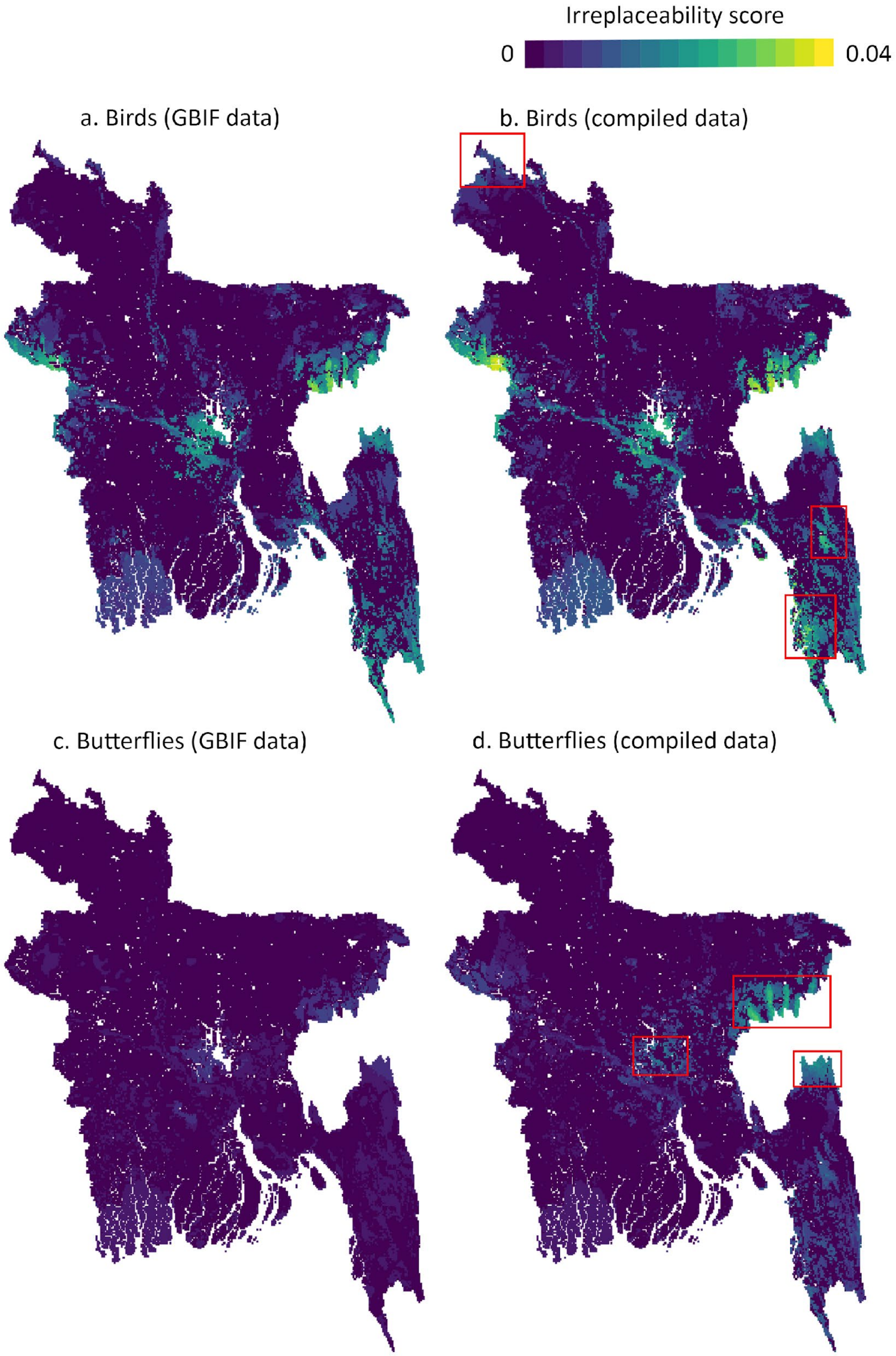

a) Land-cover data
Forests
Shrubland
Herbaceous vegetation
Herbaceous wetland
Moss and lichen
Sparse vegetation
Cropland
Built-up areas
Water bodies
b) Differences in priority cells (birds)
c) Differences in priority cells (butterflies)
Protected areas
Using compiled data
Using GBIF data
Both
None

Figure 3


Irreplaceability score
0
0.04
a. Birds (GBIF data)
b. Birds (compiled data)
c. Butterflies (GBIF data)
d. Butterflies (compiled data)

Supplementary section

Figure S1: Distribution of geospatial records for animals (birds and butterflies) in Bangladesh using records from both Facebook and the Global Biodiversity Information Facility (GBIF).

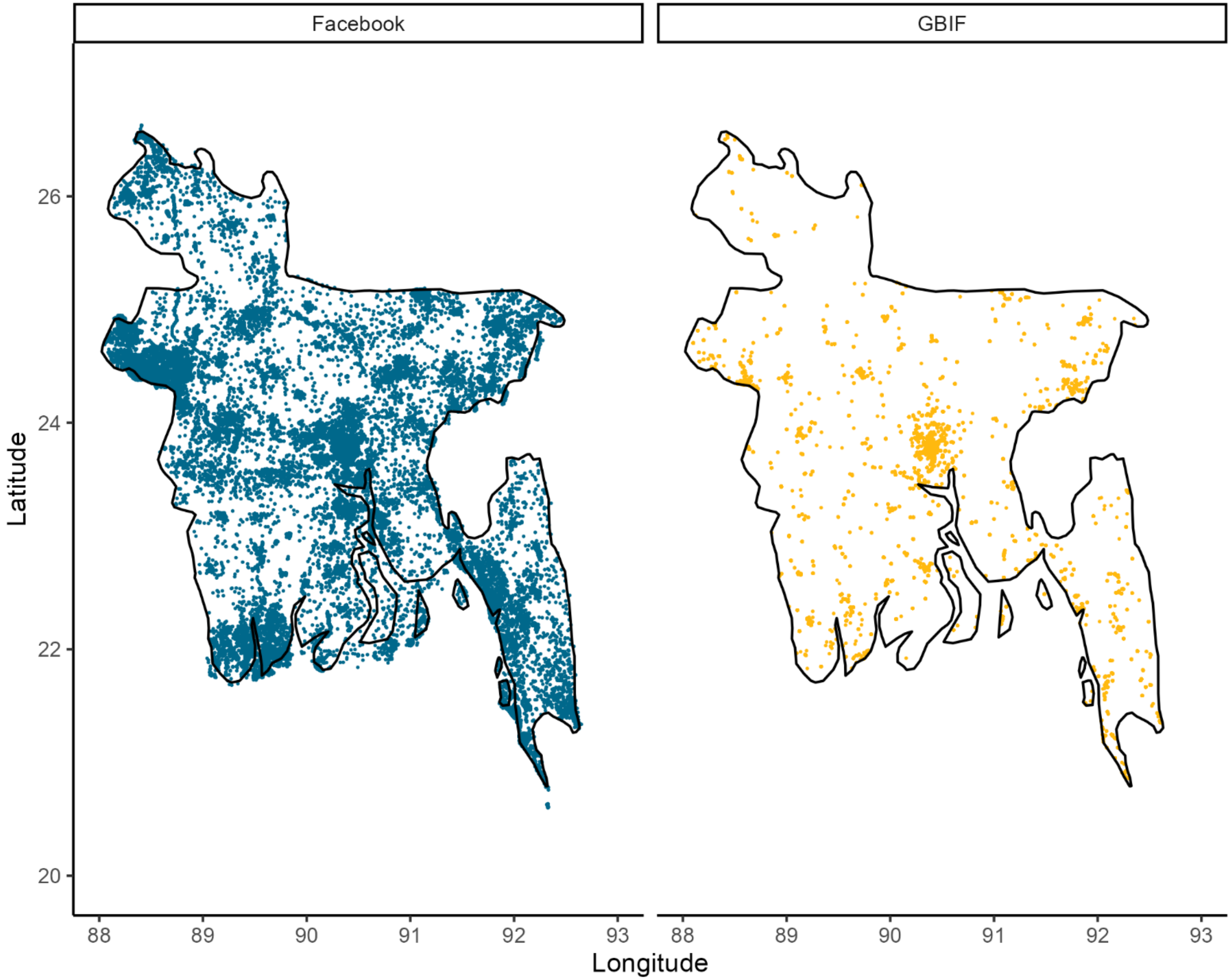


Figure S2: Maps show current protected areas and priority areas in Bangladesh for meeting species representation targets for birds and butterflies. Here, we only considered species for which we obtained suitability maps in both Global Biodiversity Information Facility (GBIF)-only and combined data approaches.

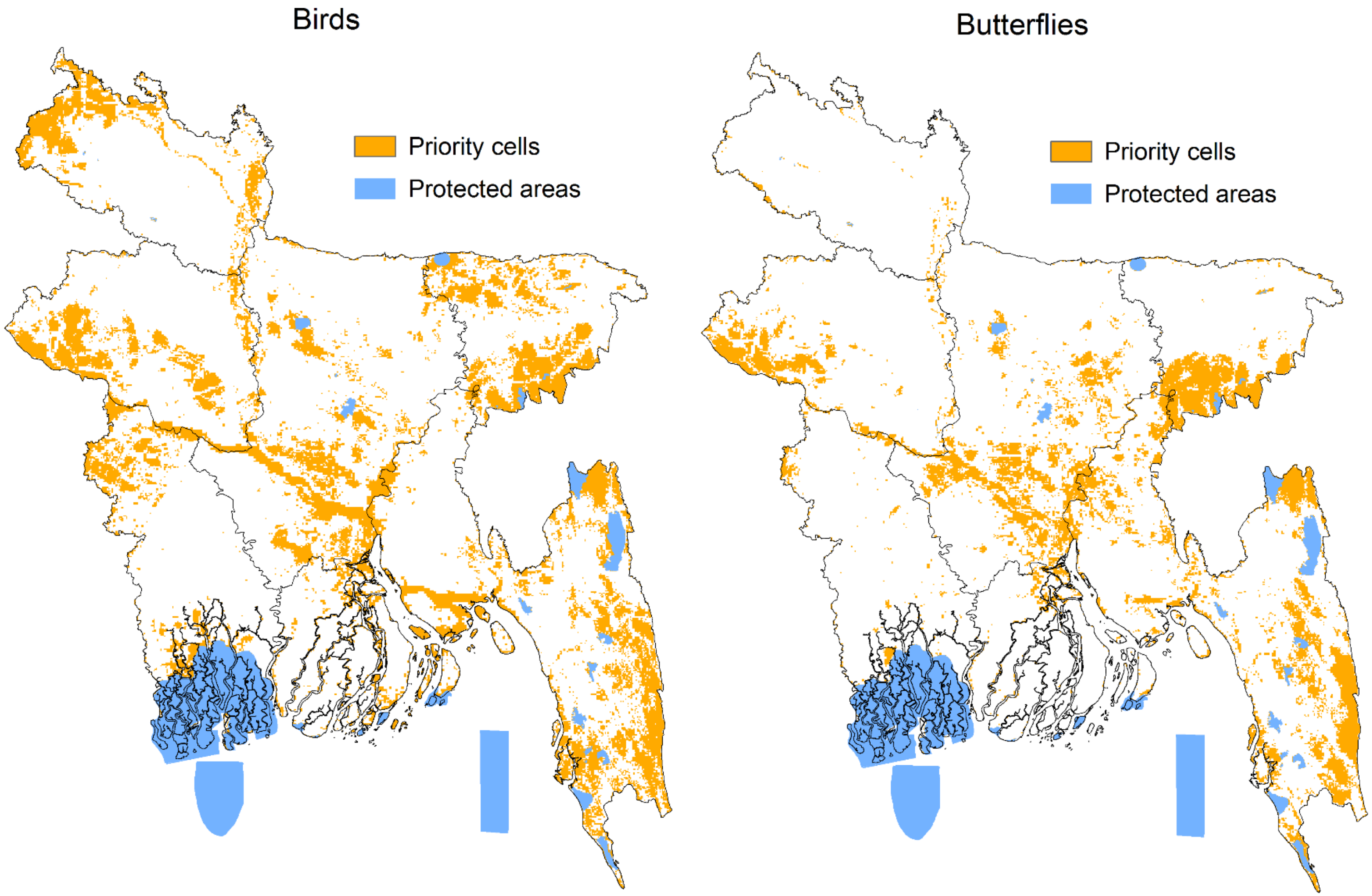


Figure S3: Histograms showing the distribution of land use within priority areas.

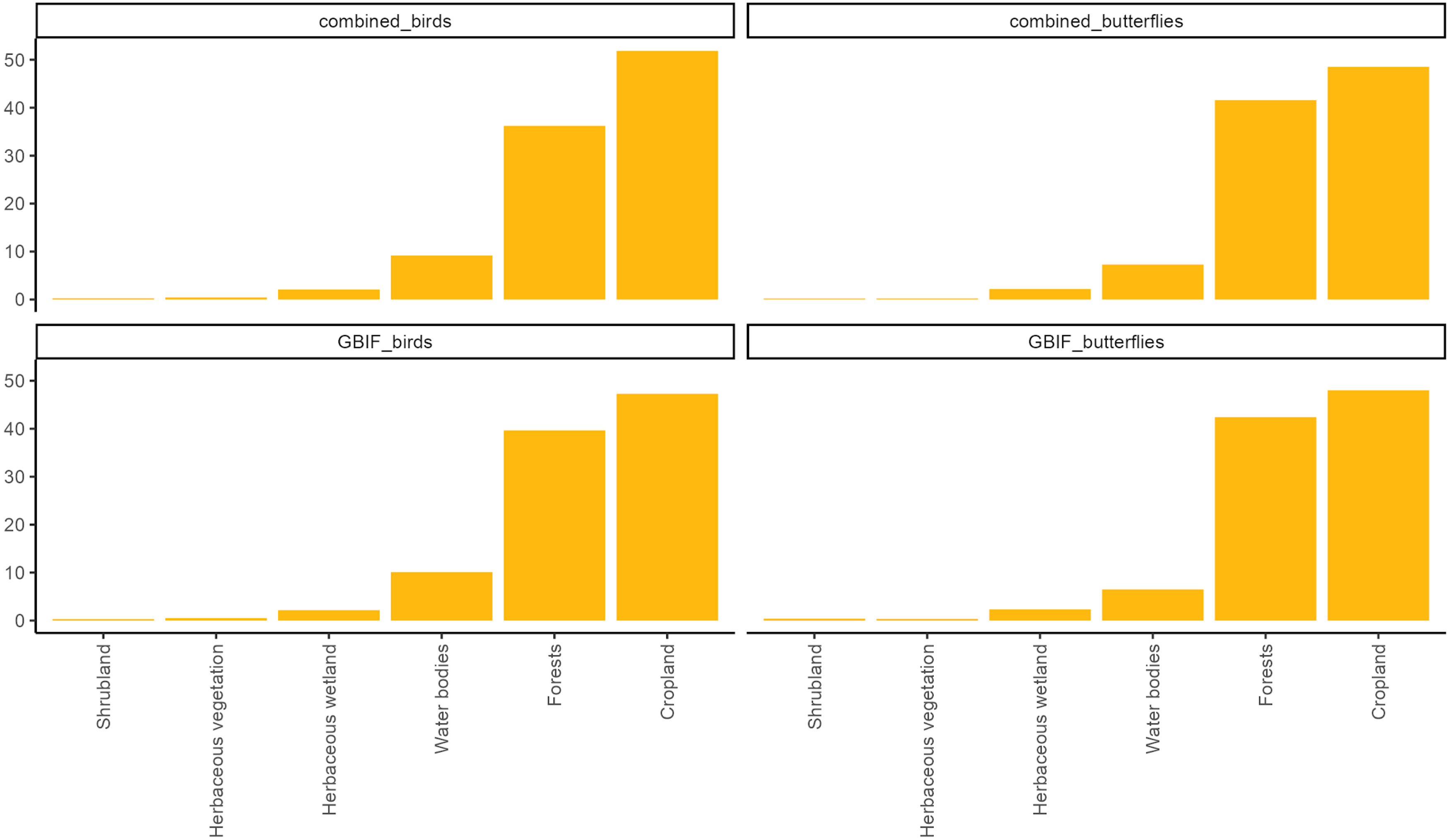


Figure S4: Histograms showing the distribution of human footprint index within priority areas.