ECOGRAPHY

*Research*

# Survey completeness of a global citizen-science database of bird occurrence

**Frank A. La Sorte and Marius Somveille**

*F. A. La Sorte (http://orcid.org/0000-0001-8521-2501) ✉ (fal42@cornell.edu), Cornell Laboratory of Ornithology, Cornell Univ., Ithaca, NY, USA. – M. Somveille (https://orcid.org/0000-0002-6868-5080), BirdLife International, The David Attenborough Building, Cambridge, UK.*

**Ecography**
**43: 34–43, 2020**
doi: 10.1111/ecog.04632

Subject Editor: Joaquin Hortal
Editor-in-Chief: Miguel Araújo
Accepted 10 September 2019

Measuring the completeness of survey inventories created by citizen-science initiatives can identify the strengths and shortfalls in our knowledge of where species occur geographically. Here, we use occurrence information from eBird to measure the survey completeness of the world's birds in this database at three temporal resolutions and four spatial resolutions across the annual cycle during the period 2002 to 2018. Approximately 84% of the earth's terrestrial surface contained bird occurrence information with the greatest concentrations occurring in North America, Europe, India, Australia and New Zealand. The largest regions with low levels of survey completeness were located in central South America, northern and central Africa, and northern Asia. Across spatial and temporal resolutions, survey completeness in regions with occurrence information was 55–74% on average, with the highest values occurring at coarser temporal and coarser spatial resolutions and during spring migration within temperate and boreal regions. Across spatial and temporal resolutions, survey completeness exceeded 90% within ca 4–14% of the earth's terrestrial surface. Survey completeness increased globally from 2002 to 2018 across all months of the year at a rate of ca 3% $yr^{-1}$. The slowest gains occurred in Africa and in montane regions, and the most rapid gains occurred in India and in tropical forests after 2012. Thus, occurrence information from a global citizen-science program for a charismatic and well-studied taxon was geographically broad but contained heterogeneous patterns of survey completeness that were strongly influenced by temporal and especially spatial resolution. Our results identify regions where the application of additional effort would address current knowledge shortfalls, and regions where the maintenance of existing effort would benefit long-term monitoring efforts. Our findings highlight the potential of citizen science initiatives to further our knowledge of where species occur across space and time, information whose applications under global change will likely increase.

Keywords: bird occurrence, citizen science, eBird, full annual cycle, survey completeness, temporal resolution

## Introduction

There are many shortfalls in our knowledge of the world's biodiversity, and one of the most basic is the Wallacean shortfall or the lack of knowledge on where species

NORDIC SOCIETY OIKOS

www.ecography.org

occur geographically (Lomolino 2004, Whittaker et al. 2005, Hortal et al. 2015, Proença et al. 2017). This deficiency has broad ramifications for research and conservation, and addressing this knowledge gap remains a significant challenge for many taxa (Whittaker et al. 2005, Boakes et al. 2010, Meyer et al. 2015). The development of citizen science programs that outsource the collection and classification of biological and ecological data (Bonney et al. 2009, Dickinson et al. 2010, Kullenberg and Kasperowski 2016) has the potential to reduce this shortfall (Troia and McManamay 2016, Chandler et al. 2017). A good example is the eBird citizen science database, an online interface where volunteers enter bird observations from any location during any time of year (Sullivan et al. 2014). The information compiled by eBird has allowed investigators to explore a broad range of questions in avian ecology, evolution and conservation (Sullivan et al. 2014, La Sorte et al. 2018), including estimating bird species distributions (Fink et al. 2010, 2014). The unique full annual cycle perspective provided by eBird has particular relevance for migratory species whose seasonal movements and environmental associations are often poorly documented (Newton 2003). An important first step in determining the ability of eBird to address the Wallacean shortfall is to quantify the completeness of the occurrence information in eBird across regions, seasons and years.

Measuring survey completeness determines the ability of survey inventories to capture the full assemblage of species that are expected to occur at a given location during a given time (Colwell and Coddington 1994). When applied to eBird occurrence information, calculating survey completeness can determine where and when this information is the most comprehensive and the most deficient (Jacobs and Zipf 2017), and how these patterns have changed over time across regions and seasons. Here, we estimate the survey completeness of eBird occurrence information globally at three temporal resolutions (day, week and month) and four spatial resolutions during the period 2002 to 2018. We expect survey completeness to be highest within regions that contain large and dedicated bird watching communities, primarily in North America and Europe and we expect survey completeness to be highest during migration when bird watching activities tend to be more intensive (Sullivan et al. 2014). In addition, at a fixed spatial resolution, we expect survey completeness to increase at coarser temporal resolutions through the influence of the species–time relationship, which describes the accumulation of new species at increasing time spans (Adler and Lauenroth 2003, Adler et al. 2005). At a fixed temporal resolution, we also expect survey completeness to increase at coarser spatial resolutions through the influence of the species–area relationship (Lomolino 2000, Soberón et al. 2007, Lobo et al. 2018). By testing these predictions, our aim is to inform research and conservation efforts by determining how survey completeness for a global citizen-science initiative is defined across regions, seasons and years.

## Material and methods

The eBird citizen-science program, initiated in 2002, contains bird obervations in checklist format where species detected by sight or sound are recorded by one or more observers during a sampling event (Sullivan et al. 2014). eBird represents a semi-structured big data resource (La Sorte et al. 2018) where volunteer observers select from a number of predefined sampling protocols and where sampling effort is determined by the observer. To date (March 2019), eBird contains roughly 44.3 million hours of sampling effort. We compiled bird occurrence information from all available eBird checklists globally during the period 2002 to 2018. The data was queried on 11 January 2019 and included all sampling protocols, all levels of sampling effort, and either designation (yes/no) for the field 'all observations reported'. We only considered observations that were identified as valid in the database, and we combined observations in grouped checklists into single checklists. A total of 33 651 642 checklists were available for analysis containing 10 387 unique species (Fig. 1a; Supplementary material Appendix 1 Fig. A1). For our analysis, we aggregated eBird checklists within equal-area hexagon cells at four spatial resolutions (49 811; 199 244; 796 977; and 3 187 910 $km^2$) of a global hexagon coverage generated using an icosahedral discrete global grid system based on a Fuller icosahedral projection using an aperture 4 hexagon partition method (Sahr et al. 2003, Sahr 2011). All analyses were conducted within the terrestrial regions of the earth, excluding marine environments and Antarctica.

We estimated survey completeness across checklists within hexagon cells using the approach described by Lobo et al. (2018). We implemented the analysis for hexagons at the finest spatial resolution (49 811 $km^2$) at three temporal resolutions (day, week and month), and we implemented the analysis for hexagons at the three coarser spatial resolutions at a daily temporal resolution. To examine patterns of survey completeness during the combined period 2002 to 2018, we formatted all occurrence information into species-by-location matrices for each combination of temporal and spatial resolution where each checklist was treated as an independent survey. To examine annual trends in survey completeness, we formatted the occurrence information by month for each individual year at the finest spatial resolution. We estimated survey completeness using the 'exact' species accumulation curve estimator (Ugland et al. 2003). This approach models the smooth relationship between the number of species and sampling effort (number of checklists). We removed poorly sampled hexagon cells from our analysis whose survey completeness estimates are likely to be of poor quality based on two criteria: the hexagon cell contained less than 10 checklists, and the ratio between the number of occurrence records and the number of observed species was $\geq 1.5$. We selected these criteria based on their ability to identify extreme outliers in our analysis (Supplementary material Appendix 1 Fig. A2).

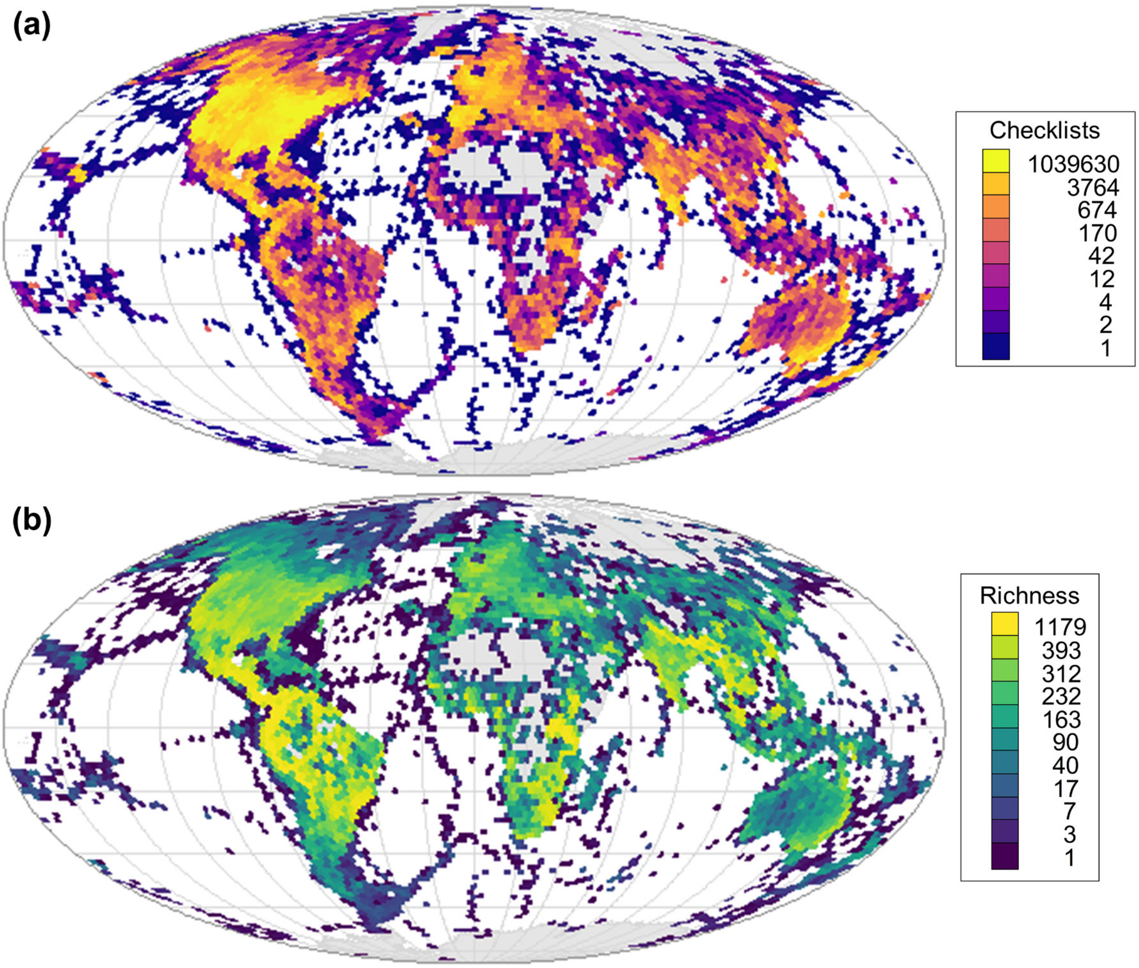


Figure 1. (a) The number of checklists in the eBird citizen-science database and (b) annual observed species richness based on occurrence information from the eBird citizen-science database compiled within equal-area hexagon cells (49 811 $km^2$) during the combined period 2002 to 2018.

We examined how survey completeness was defined across days, weeks and months during the combined period 2002 to 2018 at the finest spatial resolution using generalized additive mixed models (GAMM) (Wood 2017). We used a cyclic penalized cubic regression spline to join smoothly the first day, week, or month of the year with the last day, week, or month of the year. We included the intercept for each hexagon cell as a random effect. To provide a spatial context for interpretation, we applied GAMM using two biogeographical classifications: realms and biomes. We classified each hexagon cell within fourteen biomes (Table 1) and six biogeographical realms: Afrotropics, Australasia, IndoMalaya, Nearctic, Neotropics and Palearctic (Pielou 1979, Olson and Dinerstein 2002). At the three temporal resolutions, hexagon cells were included in the analysis if they contained estimates of survey completeness for at least 100 days, ten weeks, or four months. We also examined how survey completeness changed across years by month, realm and biome using GAMM with the intercept for hexagon cell included as a random effect. Hexagon cells were included in this analysis if they contained estimates of survey completeness for at least six years. Survey completeness was averaged across months before the GAMM was applied by realm and biome.

We determined how sampling effort affected spatial and temporal patterns of survey completeness using the following approach. We estimated the duration of sampling effort (hours) required to achieve at least 5% survey completeness for each month at the finest spatial resolution. We chose this sampling threshold and temporal resolution to allow for a more comprehensive geographic assessment. We first compiled eBird checklists by month and hexagon cell where the sampling effort was less than or equal to four hours. We selected the four hour threshold because species accumulation curves based on eBird occurrence information within individual checklists tend to reach an asymptote before the four hour mark (Kelling et al. 2015). We estimated species accumulation over time for each month and hexagon cell across checklists using generalized additive models (GAM) (Wood 2011) with the error distribution modeled as Poisson. We used the GAM fits to estimate the duration needed to

Table 1. The fourteen biomes considered in the analysis and the percent of global terrestrial surface area contained in each biome.

| ID | Biome | Area (%) |
|---|---|---|
| TSMBF | Tropical and subtropical moist broadleaf forests | 13.4 |
| TSDBF | Tropical and subtropical dry broadleaf forests | 2.6 |
| TSCF | Tropical and subtropical coniferous forests | 0.4 |
| TBMF | Temperate broadleaf and mixed forests | 8.7 |
| TCF | Temperate conifer forests | 2.0 |
| BF/T | Boreal forests/taiga | 10.9 |
| TSGSS | Tropical and subtropical grasslands, savannas and shrublands | 13.2 |
| TGSS | Temperate grasslands, savannas and shrublands | 6.5 |
| FGS | Flooded grasslands and savannas | 0.7 |
| MGS | Montane grasslands and shrublands | 3.5 |
| T | Tundra | 7.7 |
| MFWS | Mediterranean forests, woodlands and scrub | 2.2 |
| DXS | Deserts and xeric shrublands | 18.9 |

achieve 5% of expected species richness from the survey completeness analysis for each month and hexagon cell.

All analyses were conducted in R ver. 3.5.1 (R Development Core Team). The global hexagon coverages were generated using the dggridR library (Barns 2018). The survey completeness analysis was conducted using the KnowBPolygon function in the KnowBR library (Lobo et al. 2018), GAM was implemented using the mgcv library (Wood 2017), and GAMM was implemented using the gamm4 library (Wood and Scheipl 2017).

### Data deposition

Data available from the Dryad Digital Repository: <https://doi.org/10.5061/dryad.h9w0vt4d6> (La Sorte and Somveille 2019).

## Results

Approximately 84% of the earth's terrestrial surface contains bird occurrence information, with the greatest concentrations occurring in North America, Europe, India, Australia and New Zealand (Fig. 1a). The remainder of the earth's terrestrial surface contained no occurrence information, with the largest gaps occurring in central South America, central and northern Africa, and northern Asia (Fig. 1a). Observed species richness was greatest within tropical regions of the globe and declined towards the poles (Fig. 1b). Seasonal variation in observed species richness was strongest within temperate regions, primarily in the Northern Hemisphere (Supplementary material Appendix 1 Fig. A3).

Sufficient quantities of occurrence information for estimating survey completeness at the finest spatial resolution were available for ca 62% of the earth's terrestrial surface. When summarized by day at the finest spatial resolution, survey completeness was 55% on average with ca 4% of the earth's terrestrial surface containing survey completeness exceeding 90% on average (Fig. 2a). When summarized by week, survey completeness was 64% on average with ca 14% of the earth's terrestrial surface containing survey completeness exceeding 90% on average (Fig. 2c). When summarized by month, survey completeness was 72% on average with ca 14% of the earth's terrestrial surface containing survey completeness exceeding 90% on average (Fig. 2e). The primary seasonal pattern was a northward expansion of higher survey completeness within the Northern Hemisphere during the late boreal spring and early boreal summer (Supplementary material Appendix 1 Fig. A4).

Across temporal resolutions, survey completeness reached its highest levels within North America, western Europe, southern India, eastern Australia and New Zealand (Fig. 2a, c, e). At coarser temporal resolutions, survey completeness generally strengthened and expanded outside these regions, defining patterns that were globally more homogenous (Fig. 2a, c, e). Intra-annual variation in survey completeness generally declined in strength at coarser temporal resolutions, and was lowest within North America, western Europe, southern India, eastern Australia and New Zealand (Fig. 2b, d, f).

Survey completeness presented contrasting geographic patterns when summarized by day at the three coarser spatial resolution (Fig. 3). At the first spatial resolution (199 244 $km^2$), sufficient quantities of occurrence information for estimating survey completeness were available for ca 69% of the earth's terrestrial surface (Fig. 3a). Survey completeness was 58% on average with ca 5% of the earth's terrestrial surface containing survey completeness exceeding 90% on average (Fig. 3a). At the second spatial resolution (796 977 $km^2$), sufficient quantities of occurrence information for estimating survey completeness were available for ca 79% of the earth's terrestrial surface (Fig. 3c). Survey completeness was 61% on average with ca 8% of the earth's terrestrial surface containing survey completeness exceeding 90% on average (Fig. 3c). At the third spatial resolution (3 187 910 $km^2$), sufficient quantities of occurrence information for estimating survey completeness were available for ca 80% of the earth's terrestrial surface (Fig. 3e). Survey completeness was 68% on average with ca 13% of the earth's terrestrial surface containing survey completeness exceeding 90% on average (Fig. 3e).

At coarser spatial resolutions, broader geographic regions contained higher levels of survey completeness

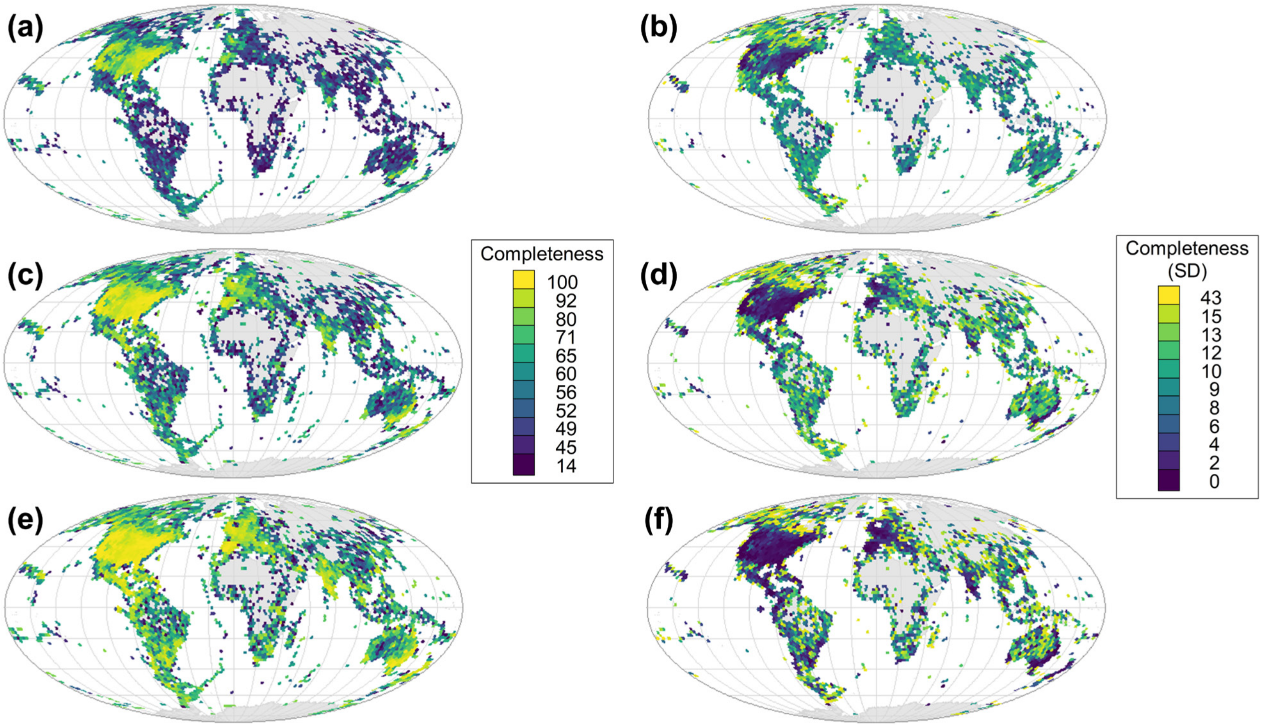


Figure 2. Survey completeness calculated and then averaged across (a) days, (c) weeks and (e) months. The standard deviation of survey completeness calculated across (b) days, (d) weeks and (f) months. Survey completeness was calculated using occurrence information from the eBird citizen-science database compiled within equal-area hexagon cells (49 811 km$^2$) during the combined period 2002 to 2018.

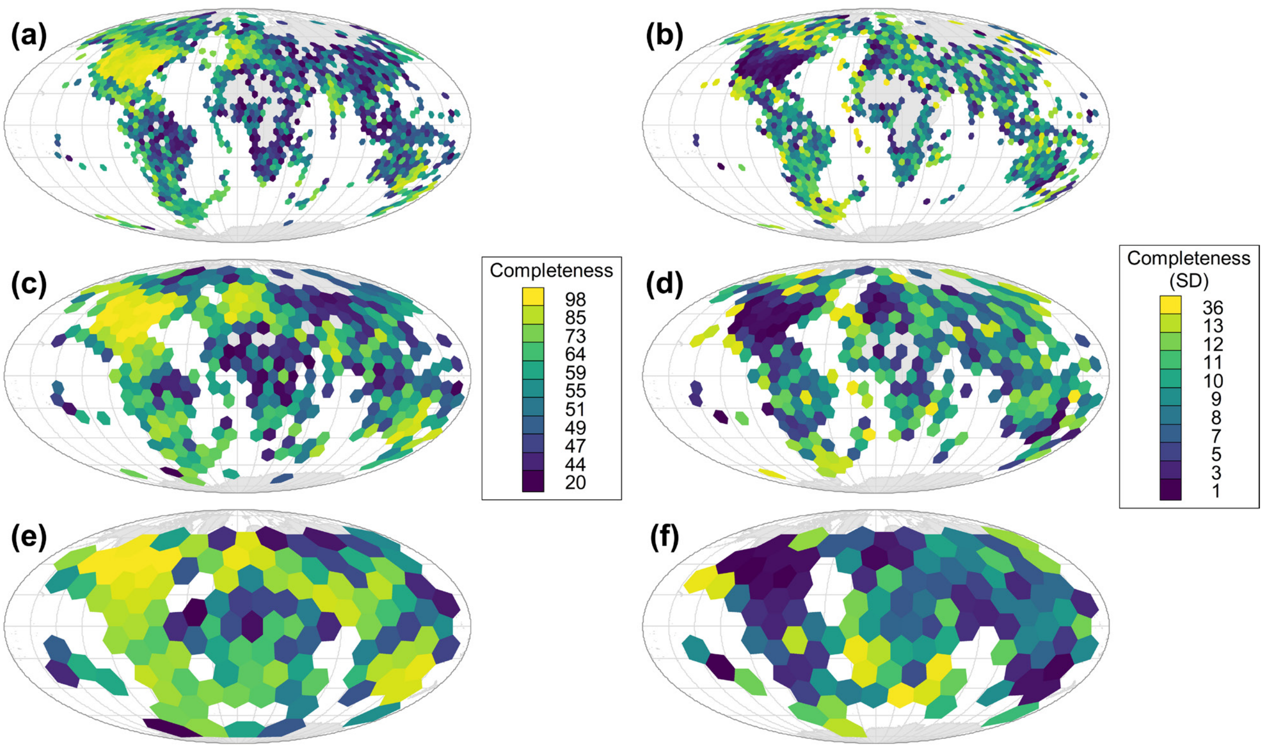


Figure 3. Survey completeness calculated and then averaged across days at three spatial resolutions (left column) and the standard deviation of survey completeness calculated across days at three spatial resolutions (right column). The three spatial resolutions include 199 244 km$^2$ (top row), 796 977 km$^2$ (middle row) and 3 187 910 km$^2$ (bottom row). Survey completeness was calculated using occurrence information from the eBird citizen-science database compiled within equal-area hexagon cells during the combined period 2002 to 2018.

(Fig. 2a, 3a, c, e). Patterns documented at the finest spatial resolution (Fig. 2a) expanded to encompass broad portions of North America, Europe, South Asia, Australia and New Zealand (Fig. 2a, 3a, c, e). The most substantial increases occurred in South America and South Africa (Fig. 2a, 3a, c, e). Intra-annual variation in survey completeness responded in a similar fashion, with the lowest values expanding to encompass broad portions of North and South America, Europe, south Asia, Australia and New Zealand, with the highest values occurring in South Africa (Fig. 2b, 3b, d, f).

When summarized by biogeographical realm, survey completeness reached its highest level on average by day (Fig. 4a), week (Fig. 4b), and month (Fig. 4c) in the Nearctic with the remaining realms presenting lower but similar levels of survey completeness. Survey completeness across all six realms tended to increase on average at coarser temporal resolutions (Fig. 4a–c). Differences among the six realms tended to decline at coarser temporal resolutions with the exception of the Afrotropics (Fig. 4a–c). When examined seasonally, survey completeness reached its highest levels across temporal resolutions in the Northern Hemisphere (Nearctic and Palearctic) during the late boreal spring and early boreal summer (Fig. 4a–c). Seasonal patterns were less variable with the four Southern Hemisphere realms; the highest levels occurred during the austral spring in Australasia and Afrotropics, and the highest levels occurred during the austral summer in Indomalaya and Neotropics (Fig. 4a–c).

When summarized by biome, survey completeness reached its highest levels on average by day (Fig. 4d), week (Fig. 4e), and month (Fig. 4f) in the three temperate biomes (TCF, TGSS and TBMF; Table 1). Survey completeness across all fourteen biomes tended to increase on average at coarser temporal resolutions, and differences among the fourteen biomes tended to decline at coarser temporal resolutions (Fig. 4d–f). When examined seasonally, survey completeness reached its highest levels during the late boreal spring and early boreal summer, with the strongest transition occurring at a daily temporal resolution within the three temperate biomes (TCF, TGSS and TBMF) and two Arctic biomes (BF/T and T; Table 1; Fig. 4d). These seasonal patterns remained strongest for the

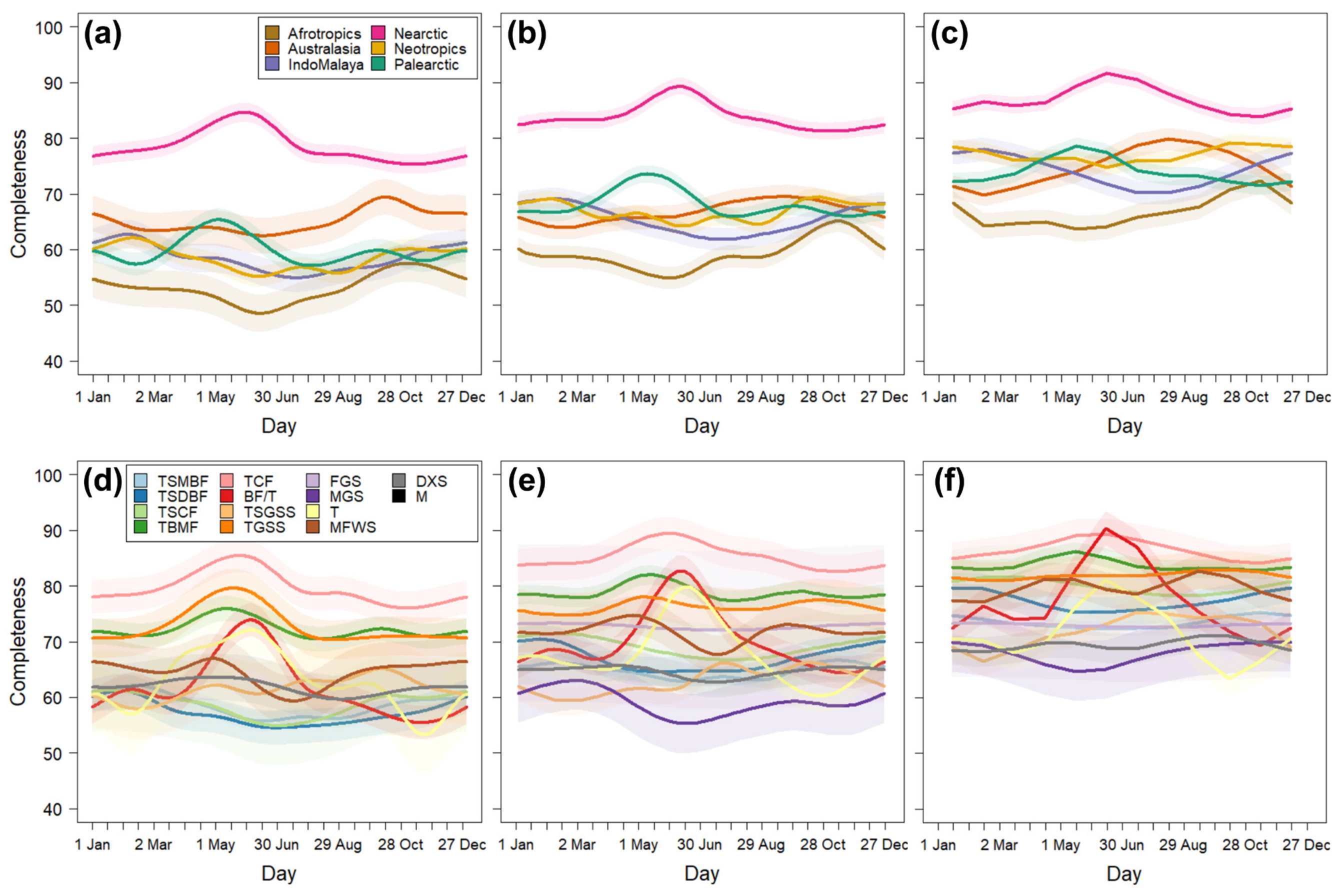


Figure 4. Survey completeness calculated by (a) day, (b) week and (c) month summarized across six biogeographical realms, and survey completeness calculated by (d) day, (e) week and (f) month summarized across fourteen biomes (Table 1). Survey completeness is based on occurrence information from the eBird citizen-science database compiled within equal-area hexagon cells (49 811 km$^2$) during the combined period 2002 to 2018. The fitted lines and 95% confidence bands are from generalized additive mixed models using a cyclic penalized cubic regression spline with the intercept for hexagon cell included as a random effect. Several biomes lacked sufficient data for model convergence: (d) FGS and MGS, and (d–f) M.

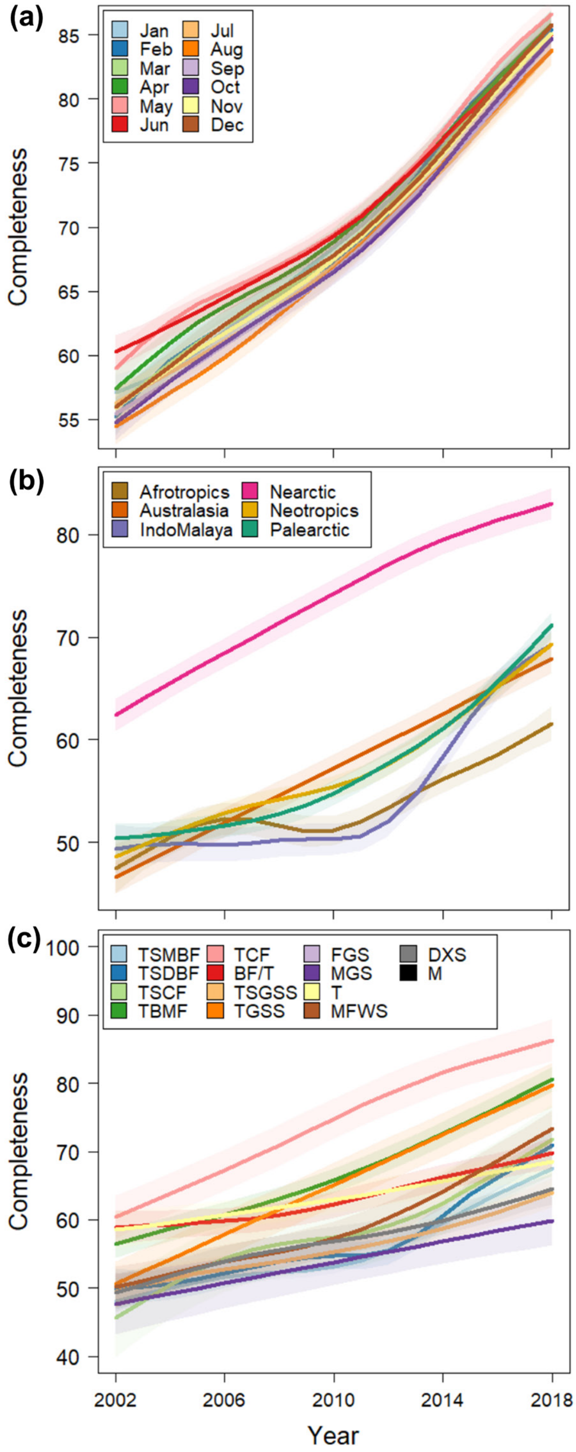


Figure 5. Survey completeness calculated by month for the years 2002 to 2018 and summarized across years by (a) month, (b) biogeographical realm and (c) biome (Table 1) using occurrence information from the eBird citizen-science database compiled within equal-area hexagon cells (49 811 km$^2$). The fitted lines and 95% confidence bands are from generalized additive mixed models with the intercept for hexagon cell included as a random effect. Two biomes (FGS and M) lacked sufficient data for model convergence.

two Arctic biomes (BF/T and T) at coarser temporal resolutions (Table 1; Fig. 4d–f). Three biomes (FGS, MGS and M) lacked sufficient data for model convergence at a daily temporal resolution, and one biome (M) lacked sufficient data for model convergence at a weekly and monthly temporal resolutions (Table 1; Fig. 4d–f).

When examined across years, survey completeness increased from 2002 to 2018 in a similar linear fashion for each month of the year (Fig. 5a). These trends generated an average increase of ca 51% in survey completeness for each month from 2002 to 2018 (average annual rate ca 3%). Survey completeness increased in a more variable fashion from 2002 to 2018 when examined by biogeographical realm (Fig. 5b). The Nearctic had the largest overall levels of survey completeness on average, starting at ca 62 in 2002 and ending at ca 82 in 2018 (Fig. 5b). Survey completeness for all the remaining realms started on average between ca 45 and 50 in 2002 and ended at ca 65, except for the Afrotropics which ended below 60 (Fig. 5b). The slowest gains occurred in Africa and the most rapid gains occurred in India after 2012 (Fig. 5b). Survey completeness increased from 2002 to 2018 in a more variable fashion when examined by biome (Table 1; Fig. 5c). The strongest gains occurred with the three temperate biomes (TBMF, TCF and TGSS; Table 1; Fig. 5c). Five biomes showed weak overall gains (BF/T, TSGSS, MGS, T and DXS), three tropical forest biomes (TSMBF, TSDBF and TSCF) and the Mediterranean biome (MFWS) showed strong gains after 2012, and two biomes (FGS and M) lacked sufficient data for model convergence (Table 1; Fig. 5c).

The duration of sampling effort required to achieve 5% survey completeness across checklists based on expected species richness (Supplementary material Appendix 1 Fig. A5) was greatest on average within tropical regions and lowest within temperate regions (Fig. 6a). Seasonal variation in the duration of sampling effort required to achieve 5% survey completeness was lowest within North America, western Europe, southern India, eastern Australia and New Zealand (Fig. 6b; Supplementary material Appendix 1 Fig. A6).

## Discussion

Our global assessment of the completeness of eBird occurrence information identified the presence of strong geographic, seasonal and yearly patterns. As expected, occurrence information was concentrated in regions of North America and Europe where birdwatching activities are currently the most intensive (Sullivan et al. 2014). Secondary concentrations occurred in southern India, eastern Australia and New Zealand. These regions in combination contain the majority of world's citizen science initiatives, where crowdsourcing has been successfully used to collect biological and ecological data for birds and other taxa (Chandler et al. 2017). The highest levels of survey completeness occurred in these same regions, with North America having the highest overall levels. As expected, survey completeness was higher during spring migration within temperate and boreal regions of

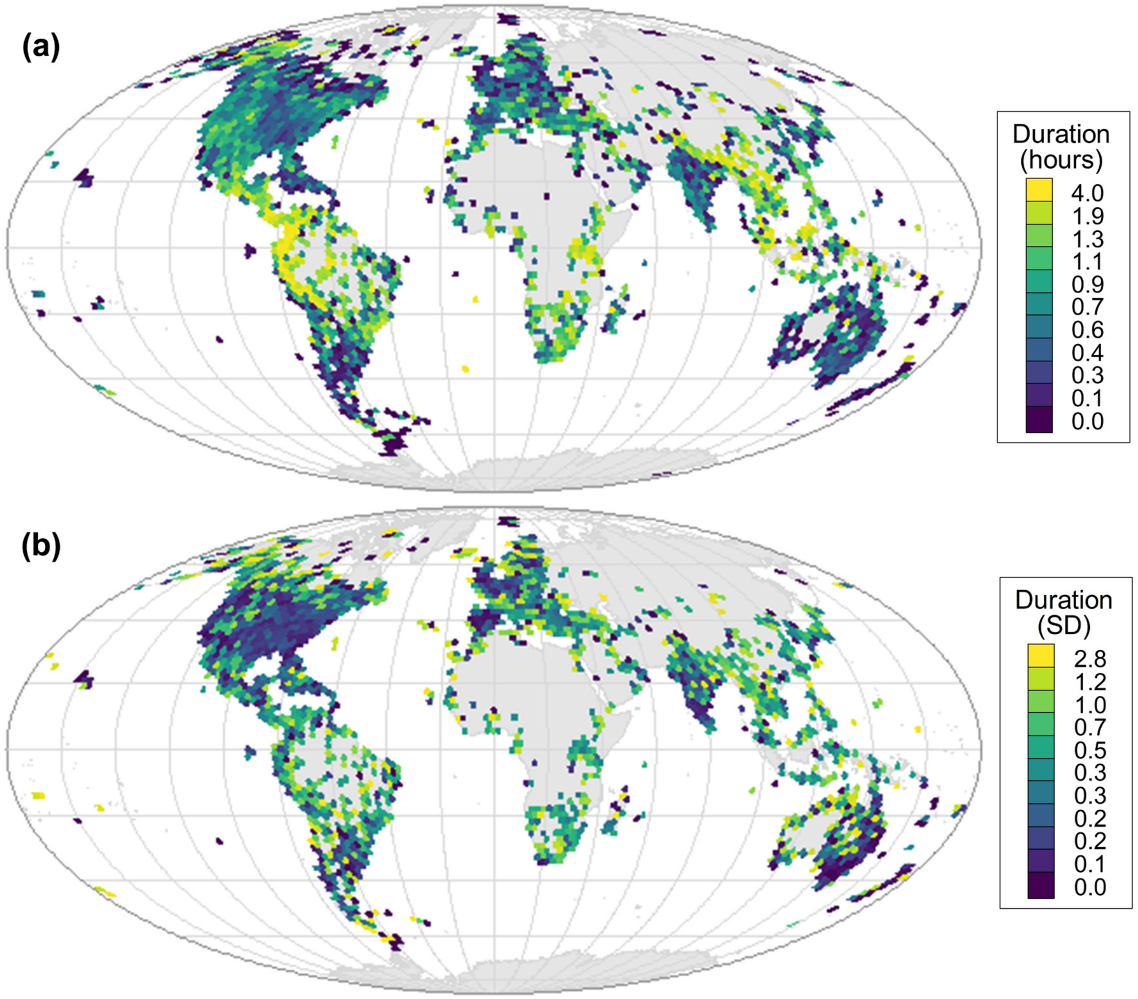


Figure 6. (a) The duration of survey effort (hours) needed to achieve 5% survey completeness averaged across months, and (b) the standard deviation across months of the duration of survey effort needed to achieve 5% survey completeness. Survey completeness was calculated using occurrence information from the eBird citizen-science database compiled monthly within equal-area hexagon cells (49 811 $km^2$) during the combined period 2002 to 2018.

North America and Europe when birdwatching activities tend to be more intensive. Large regions in South America, Africa, and northern Asia contained no occurrence information or low levels of survey completeness. These same regions have been identified as containing significant gaps in occurrence information for other taxa (Meyer et al. 2015). Survey completeness increased from 2002 to 2018 in a consistent fashion across seasons, with the strongest gains occurring in India and tropical forests after 2012 and the weakest gains occurring in Africa and montane regions. The quantity of sampling effort required to achieve a minimum level of survey completeness was greatest in tropical regions of the globe where species richness tends to be highest, and seasonal variation in this effort tended to be the lowest in the regions listed above that have the highest levels of survey completeness.

Our analysis provided the first consideration of the effect of temporal resolution on survey completeness. In agreement with the species–time relationship (Adler and Lauenroth 2003, Adler et al. 2005), our findings indicate that, at a fixed spatial resolution, coarser temporal resolutions contained more complete bird occurrence information, and this was the case for all regions of the globe. Thus, the accumulation of occurrence information through the coarsening of temporal resolutions increased survey completeness. Within species-rich tropical regions of the globe, this increase is likely related to the addition of occurrence information on rare species, which are more prevalent in these regions (Jenkins et al. 2013). Within species-poor temperate regions of the globe, this increase is likely related to the addition of occurrences information on transient species that are observed during migration and occur in greater proportions in these regions (Somveille et al. 2013). When examined within the context of sampling effort, additional effort was required in tropical regions to match the level of survey completeness obtained in temperate regions, a likely outcome of the greater number of species in the tropics and the greater prevalence of rare species (Jenkins et al. 2013).

Following our expectations, survey completeness increased at coarser spatial resolutions (Soberón et al. 2007, Lobo et al. 2018). This outcome resulted in survey completeness increasing across broad geographic regions, in some cases encompassed regions that lacked occurrence information or contained low levels of survey completeness. For example, coarser spatial resolution increased survey completeness across the full extent of the South American continent. Thus, if the available checklists offer a good representation of the environmental and compositional heterogeneity within a region (Lobo et al. 2018), coarsening the spatial resolution has the potential to generate more complete estimates of species composition in poorly sampled regions of the globe.

A primary application of occurrence information is with species distribution models (SDMs) where it is used to extract environmental associations across a species' distribution (Elith and Leathwick 2009). The quality of the occurrence information directly determines SDM performance (Fei and Yu 2016), and measuring survey completeness helps identify where SDMs are likely to be more or less reliable (Hortal et al. 2008, 2015). When using presence-only SDMs (Tsoar et al. 2007), estimates of species' distributions are likely to be more reliable in well surveyed regions and seasons where presence information is geographically comprehensive and the form and extent of a species' ecological niche is well represented within the environmental space (Fei and Yu 2016). Not considering survey completeness in these models could introduce spatial biases due to spatial variation in data quality and quantity. With the addition of absence information, outside of well surveyed regions and seasons, SDMs may generate less reliable estimates (Lahoz-Monfort et al. 2013, Guélat and Kéry 2018) or misleading measures of model performance (Lobo et al. 2010, Leroy et al. 2018) due to the greater likelihood of false absences. One approach to accommodate survey completeness in SDMs is to include covariates that capture aspects of the observation process, such as effort and detectability (Johnston et al. 2019, Kelling et al. 2019). If information on the observation process is sparse, survey completeness could act as a proxy in SDMs for total sampling effort.

When examining eBird occurrence information globally, our findings indicate that, within a fixed temporal window, the information is geographically broad but the number and size of well-surveyed regions is limited. However, we found that the extent of well-surveyed regions can increase substantially at coarser spatial resolutions. To enhance the potential of citizen science initiatives to fill shortfalls in our knowledge of the world's biodiversity and advance research and conservation, additional work is needed to expand survey effort within the poorly sampled regions of the globe (Pocock et al. 2018). Our findings show that some initiatives are realizing this potential through consistent gains in survey completeness across large geographic regions. To promote the use of citizen science data for long-term biodiversity monitoring, it is necessary to maintain high levels of survey effort in well-sampled regions (Pocock et al. 2018). This is particularly relevant when assessing the implications of global change for natural systems where information on where species occur can play a critical role (Chandler et al. 2017, Schmeller et al. 2017, Pocock et al. 2018). When considering species whose distributions are seasonally dynamic, such as migratory birds (La Sorte et al. 2017), the information generated by year-round citizen science programs are particularly valuable (Ådahl et al. 2006, Marra et al. 2015). Therefore, citizen science initiatives provide unique opportunities to advance biological knowledge across space and time. As the information compiled by these initiatives continues to expand in breadth and depth, its value for research and conservation is likely to improve.

*Acknowledgements* – We thank D. Fink and A. Johnston for valuable comments on an earlier draft, two anonymous reviewers for constructive suggestions, S. Kelling and the eBird team for their support, and the many eBird participants for their contributions.
*Funding* – This work was funded by The Wolf Creek Charitable Foundation and the National Science Foundation (ABI sustaining: DBI-1356308; ABI innovation: DBI-1661329; computing support from CNS-1059284 and CCF-1522054).

## References

Ådahl, E. et al. 2006. From climate change to population change: the need to consider annual life cycles. – Global Change Biol. 12: 1627–1633.

Adler, P. B. and Lauenroth, W. K. 2003. The power of time: spatiotemporal scaling of species diversity. – Ecol. Lett. 6: 749–756.

Adler, P. B. et al. 2005. Evidence for a general species–time–area relationship. – Ecology 86: 2032–2039.

Barns, R. 2018. dggridR: discrete global grids. – R package ver. 2.0.3, <https://CRAN.R-project.org/package=dggridR>.

Boakes, E. H. et al. 2010. Distorted views of biodiversity: spatial and temporal bias in species occurrence data. – PLoS Biol. 8: e1000385.

Bonney, R. et al. 2009. Citizen science: a developing tool for expanding science knowledge and scientific literacy. – Bioscience 59: 977–984.

Chandler, M. et al. 2017. Contribution of citizen science towards international biodiversity monitoring. – Biol. Conserv. 213: 280–294.

Colwell, R. K. and Coddington, J. A. 1994. Estimating terrestrial biodiversity through extrapolation. – Phil. Trans. R. Soc. B 345: 101–118.

Dickinson, J. L. et al. 2010. Citizen science as an ecological research tool: challenges and benefits. – Annu. Rev. Ecol. Evol. Syst. 41: 149–172.

Elith, J. and Leathwick, J. R. 2009. Species distribution models: ecological explanation and prediction across space and time. – Annu. Rev. Ecol. Evol. Syst. 40: 677–697.

Fei, S. and Yu, F. 2016. Quality of presence data determines species distribution model performance: a novel index to evaluate data quality. – Landscape Ecol. 31: 31–42.

Fink, D. et al. 2010. Spatiotemporal exploratory models for broad-scale survey data. – Ecol. Appl. 20: 2131–2147.

Fink, D. et al. 2014. Crowdsourcing meets ecology: hemisphere-wide spatiotemporal species distribution models. – AI Mag. 35: 19–30.

Guélat, J. and Kéry, M. 2018. Effects of spatial autocorrelation and imperfect detection on species distribution models. – Methods Ecol. Evol. 9: 1614–1625.
Hortal, J. et al. 2008. Historical bias in biodiversity inventories affects the observed environmental niche of the species. – Oikos 117: 847–858.
Hortal, J. et al. 2015. Seven shortfalls that beset large-scale knowledge of biodiversity. – Annu. Rev. Ecol. Evol. Syst. 46: 523–549.
Jacobs, C. and Zipf, A. 2017. Completeness of citizen science biodiversity data from a volunteered geographic information perspective. – Geo-spatial Inform. Sci. 20: 3–13.
Jenkins, C. N. et al. 2013. Global patterns of terrestrial vertebrate diversity and conservation. – Proc. Natl Acad. Sci. USA 110: E2602–E2610.
Johnston, A. et al. 2019. Best practices for making reliable inferences from citizen science data: case study using eBird to estimate species distributions. – bioRxiv 574392.
Kelling, S. et al. 2015. Can observation skills of citizen scientists be estimated using species cccumulation curves? – PLoS One 10: e0139600.
Kelling, S. et al. 2019. Using semistructured surveys to improve citizen science data for monitoring biodiversity. – Bioscience 69: 170–179.
Kullenberg, C. and Kasperowski, D. 2016. What is citizen science? – A scientometric meta-analysis. – PLoS One 11: e0147152.
Lahoz-Monfort, J. J. et al. 2013. Imperfect detection impacts the performance of species distribution models. – Global Ecol. Biogeogr. 23: 504–515.
La Sorte, F. A. and Somveille, M. 2019. Data from: Survey completeness of a global citizen-science database of bird occurrence. – Dryad Digital Repository, <https://doi.org/10.5061/dryad.h9w0vt4d6>.
La Sorte, F. A. et al. 2017. Global change and the distributional dynamics of migratory bird populations wintering in Central America. – Global Change Biol. 23: 5284–5296.
La Sorte, F. A. et al. 2018. Opportunities and challenges for big data ornithology. – Condor 120: 414–426.
Leroy, B. et al. 2018. Without quality presence–absence data, discrimination metrics such as TSS can be misleading measures of model performance. – J. Biogeogr. 45: 1994–2002.
Lobo, J. M. et al. 2010. The uncertain nature of absences and their importance in species distribution modelling. – Ecography 33: 103–114.
Lobo, J. M. et al. 2018. KnowBR: an application to map the geographical variation of survey effort and identify well-surveyed areas from biodiversity databases. – Ecol. Indic. 91: 241–248.
Lomolino, M. V. 2000. Ecology's most general, yet protean pattern: the species–area relationship. – J. Biogeogr. 27: 17–26.
Lomolino, M. V. 2004. Conservation biogeography. – In: Lomolino, M. V. and Heaney, L. R. (eds), Frontiers of biogeography; new directions in the geography of nature. Sinauer Associates, pp. 293–296.
Marra, P. P. et al. 2015. A call for full annual cycle research in animal ecology. – Biol. Lett. 11: 20150552.
Meyer, C. et al. 2015. Global priorities for an effective information basis of biodiversity distributions. – Nat. Commun. 6: 8221.
Newton, I. 2003. Geographical patterns in bird migration. – In: Berthold, P. et al. (eds), Avian migration. Springer, pp. 211–224.
Olson, D. M. and Dinerstein, E. 2002. The Global 200: priority ecoregions for global conservation. – Ann. Mo. Bot. Gard. 89: 199–224.
Pielou, E. C. 1979. Biogeography. – Wiley.
Pocock, M. J. O. et al. 2018. A vision for global biodiversity monitoring with citizen science. – In: Bohan, D. A. et al. (eds), Advances in ecological research. Academic Press, pp. 169–223.
Proença, V. et al. 2017. Global biodiversity monitoring: from data sources to essential biodiversity variables. – Biol. Conserv. 213: 256–263.
Sahr, K. 2011. Hexagonal discrete global grid systems for geospatial computing. – Arch. Photogr. Cartogr. Remote Sens. 22: 363–376.
Sahr, K. et al. 2003. Geodesic discrete global grid systems. – Cartogr. Geogr. Inform. Sci. 30: 121–134.
Schmeller, D. S. et al. 2017. Building capacity in biodiversity monitoring at the global scale. – Biodivers. Conserv. 26: 2765–2790.
Soberón, J. et al. 2007. Assessing completeness of biodiversity databases at different spatial scales. – Ecography 30: 152–160.
Somveille, M. et al. 2013. Mapping global diversity patterns for migratory birds. – PLoS One 8: e70907.
Sullivan, B. L. et al. 2014. The eBird enterprise: an integrated approach to development and application of citizen science. – Biol. Conserv. 169: 31–40.
Troia, M. J. and McManamay, R. A. 2016. Filling in the GAPS: evaluating completeness and coverage of open-access biodiversity databases in the United States. – Ecol. Evol. 6: 4654–4669.
Tsoar, A. et al. 2007. A comparative evaluation of presence-only methods for modelling species distribution. – Divers. Distrib. 13: 397–405.
Ugland, K. I. et al. 2003. The species–accumulation curve and estimation of species richness. – J. Anim. Ecol. 72: 888–897.
Whittaker, R. J. et al. 2005. Conservation biogeography: assessment and prospect. – Divers. Distrib. 11: 3–23.
Wood, S. N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. – J. R. Stat. Soc. B 73: 3–36.
Wood, S. N. 2017. Generalized additive models: an introduction with R. – Chapman and Hall/CRC.
Wood, S. N. and Scheipl, F. 2017. gamm4: generalized additive mixed models using mgcv and lme4. – R package ver. 0.2–5, <https://CRAN.R-project.org/package=gamm4>.

Supplementary material (Appendix ECOG-04632 at <www.ecography.org/appendix/ecog-04632>). Appendix 1.