Biological Conservation 221 (2018) 151–159

Contents lists available at ScienceDirect

Biological Conservation

journal homepage: www.elsevier.com/locate/biocon

Using opportunistic citizen science data to estimate avian population trends

Joshua J. Horns[a,*], Frederick R. Adler[a,b], Çağan H. Şekercioğlu[a,c]

[a] University of Utah, Department of Biology, Salt Lake City, UT 84112, USA
[b] University of Utah, Department of Mathematics, Salt Lake City, UT 84112, USA
[c] College of Sciences, Koç University, Rumelifeneri, Istanbul 34450, Turkey

ARTICLE INFO

Keywords:
Avian ecology
Biodiversity monitoring
Birding
Breeding bird survey
Conservation biology
eBird
List length analysis
Ornithology
Population trend
USGS

ABSTRACT

Determining population trends is critical for conservation. For most bird species, trends are based on count data gathered by institutions with formalized survey protocols. However, limited resources may prevent these types of surveys, especially in developing countries. Ecotourism growth and subsequent increases in opportunistic data from birdwatching can provide a source of population trend information if analyses control for inter-observer variation. List length analysis (LLA) controls for such variation by using the number of species recorded as a proxy for observer skill and effort. Here, we use LLA on opportunistic data gathered by eBird to estimate population trends for 574 North American bird species (48% of species declining) and compare these estimates to population trends based on 1) formal breeding bird surveys (54% of species declining) and 2) population estimates from eBird data controlled using more rigorous correction (46% of species declining). Our analyses show that eBird data produce population trends that differ on average by only 0.4%/year from formal surveys and do not differ significantly from estimates using more control metrics. We find that estimates do not improve appreciably beyond 10,000 checklists, suggesting this as the minimum threshold of opportunistic data required for population trend estimation. Lastly, we show that characteristics affecting a species' ubiquity, such as geographic and elevational range, can affect its population trend estimate. Our results suggest that opportunistic data can be used to approximate species population trends, especially for widespread species. Because our protocol uses information present in all checklists, it can be applied to a diversity of data sources including eBird, trip reports, and bird atlases.

1. Introduction

Determining regional and large-scale population trends for species is a critical component of conservation. Accurate population trends are required to identify species of conservation concern and to evaluate the effectiveness of conservation programs (Kleiman et al., 2000; Tear et al., 1995). For most bird species, populations are monitored using point count data (Howe et al., 1989; Robertson et al., 1995; Sauer et al., 2017), which assumes that changes in how often a species is detected are correlated with changes in that species overall abundance.

In North America, the United States Geological Survey (USGS) and the Canadian Wildlife Service (CWS) oversee the annual North American Breeding Bird Survey (BBS) to monitor the populations of many bird species that are breeding residents (Sauer et al., 2017). The BBS maintains thousands of transects where observers record all birds detected visually or aurally at set locations. These counts generate reliable population trends for many bird species at the state and national level (Downes et al., 2016). Monitoring programs such as this require substantial resources and are absent from most developing countries (Seak et al., 2012; Şekercioğlu, 2012a), even though the growing threat of climate change has made such monitoring programs more important than ever (Harris et al., 2011). This paucity of population monitoring is especially true in the tropics, home to the majority of the world's bird species, many of which are specialized, sedentary and threatened with extinction (Şekercioğlu and Sodhi, 2007; Tobias et al., 2013). Only a few tropical and/or developing countries have bird atlas data (Robertson et al., 1995) while in most countries ornithological data primarily come from birdwatching tours, individual birdwatchers, and other forms of opportunistic data (Şekercioğlu, 2012a). The geographical and temporal coverage of these types of data are less systematic than those of the BBS and may result in less accurate population estimates.

The increase in ecotourism and the development of large citizen science programs have resulted in a rapidly growing body of data on birds. Opportunistic data have been previously employed to effectively answer questions about species occurrence at large geographic or temporal scales (Devictor et al., 2010). In some studies, large volumes of opportunistic data have yielded results similar to those of formal

---

[*] Corresponding author.
E-mail address: joshua.horns@utah.edu (J.J. Horns).

https://doi.org/10.1016/j.biocon.2018.02.027
Received 25 October 2017; Received in revised form 18 February 2018; Accepted 21 February 2018

bird-count surveys when examining spatial or temporal patterns of bird occurrence (Munson et al., 2010; Walker and Taylor, 2017). Others, however, have cautioned against the use of opportunistic data, particularly when estimating population trends for common species (Kamp et al., 2016).

eBird is a large citizen science database that contains a large and growing volume of bird count data (hereafter "checklists" or "lists") (Sullivan et al., 2009). Data from eBird has been successfully used to analyze diversity (Callaghan and Gawlik, 2015; La Sorte et al., 2014), species distributions (Fink et al., 2010), and migration (Supp et al., 2015), as well as monitor population trends (Clark, 2017; Walker and Taylor, 2017). These data are submitted by participants with a wide range of skill and experience, and thus some means of observer quality control must be implemented in any analysis. All eBird checklists are submitted with various metrics that can help control for variation among observers. Each checklist has data on the number of observers, the time spent observing, and the distance travelled. This information plays an important role in standardizing observations across participants, but is not available in many data sources, such as birding tour lists and bird atlas data.

List length analysis (LLA) uses only the number of species recorded on a given list as a proxy for observer skill and effort (Szabo et al., 2010). LLA operates under the assumption that as the number of species recorded on a given checklist increases, the likelihood of that list recording a specific species also increases. Previous analyses of eBird data have confirmed that the number of species reported increases with both time spent observing (effort) and long-term participation continuity (skill) (Kelling et al., 2015). Studies of eBird data have also shown that using the number of species recorded does help to control for inter-observer variability when estimating occupancy (Johnston et al., 2017) Because the number of species observed can be gathered from any birdwatching checklist, the use of LLA would allow for data from a greater number of sources to be used when estimating population trends.

However, using LLA in place of eBird's more complete means of quality control may produce unreliable trend estimates. Population trend estimates can vary widely depending on the method of analysis used (Thomas and Martin, 1996). It has also been suggested that LLA may perform poorly in areas with low diversity (Isaac et al., 2014). Therefore, we use two different methods to estimate bird population trends from opportunistic data (eBird) and compare them with each other and with estimates from more formal surveys (BBS). Our first method of analysis, hereafter "additional parameters," or "AP," uses multiple parameters of effort associated with each eBird checklists, including distance travelled and time spent observing. Our second method, hereafter "list-length-only," or "LO" uses only LLA, testing its ability to serve as a proxy for effort. LLA has previously been used as a means of quality control with eBird data (Walker and Taylor, 2017), though only in conjunction with other metrics. Here, we compare results generated from more complex models to those generated by models that only use LLA as a means of quality control. If results from each analytical method are similar, it may be feasible to use multiple sources of opportunistic data (such as birding tour lists and bird atlas information) for which more standard methods of quality control may be unavailable.

In this paper, we compare avian population trend data gathered by formal surveying (BBS) to those estimated using LLA and eBird data. We estimate population trends from eBird using both AP and LO analytical methods and compare these methods to one another. We also test the ability of citizen science data to estimate overall population trajectories (the proportion of species with increasing or decreasing trends) at a broad regional scale. We then use these results to estimate the volume of citizen science data required to accurately detect these large-scale changes. Finally, we investigate avian ecological characteristics that best predict the potential of a species' population to be reliably estimated using this methodology.

2. Methods

2.1. Data selection and trend calculation

We analyzed population changes for 574 bird species that occur on both the Breeding Bird Survey lists and eBird checklists. All analyses were done using R (Version 3.1.1) (R Core Team, 2014).

2.1.1. BBS trends

We downloaded the complete BBS dataset and reduced it to records from the contiguous 48 United States (Paradiek et al., 2017). We further reduced the dataset to counts conducted from 1997 through 2016. The BBS dataset contains records as far back as 1967, however before 1997 most years contain fewer than 100 records and no years prior to 1997 contain > 10,000. Starting in 1997 all years contain between 128,000 and 141,000 records. Species that were recorded to the sub-species level by the BBS were lumped together. We then generated presence/absence data for each species at each point count station. Analyses were done using presence/absence data rather than abundance to make the results comparable between the BBS and eBird because many eBird lists do not report abundance. Previous studies have found strong linear correlations between the proportion of BBS point count stations at which a species occurred and the reported abundance (Walker and Taylor, 2017). Species population trends were estimated by fitting their presence/absence data to mixed logistic regression models, with year treated as a fixed effect. To reduce error associated with geographic variation, route ID nested within state was treated as a random effect. To ensure that using presence/absence data in place of abundance data did not seriously affect trend estimates, we re-calculated population trends by using BBS abundance data and mixed Poisson regression. The rest of the cofactors from the logistic regression were kept the same. The Pearson correlation coefficient across all species was 0.74, suggesting a high degree of correlation between presence/absence and abundance-based modeling techniques.

2.1.2. eBird trends

We downloaded the complete eBird basic dataset and again reduced it to checklists from the contiguous United States gathered between 1997 and 2016. Checklists were based on unique "sampling event identifiers." eBird users are required to specify if they are reporting all birds detected or whether their list represents only a sample of the present avifauna. We eliminated all checklists that users defined as incomplete. We also eliminated any checklists with fewer than four species, as these short lists often represent a targeted search for a specific species and have the potential to confound results (Szabo et al., 2010). Duplicate lists were excluded by condensing lists on the basis of "group identifier". 11,681,254 eBird lists remained for analysis after duplicate, incomplete, and short lists were removed. When estimating population trends for each species, we only used checklists from eBird locations, as defined by the "locality ID", with at least one record for that species. All checklists that met the criteria for analysis were assigned a 1 or 0 depending on whether they recorded the species of interest. We generated two sets of population trends for each species by fitting this presence/absence data from eBird checklists to either an AP or LO multiple logistic model. Both models included "year", "number of species", and "state" as fixed effects. Every species observed during one observation period receives its own record in eBird's data, but all are associated with the same "sampling event identifier". Therefore we determined number of species recorded as the number of times a unique "sampling event identifier" appeared in the data. AP models also took advantage of the metrics of quality control associated with all eBird checklists by including "distance travelled", and "time spent observing" as fixed effects. We additionally ran the same models but only used the most recent 5, 10, and 15 years of data from eBird (rather than the 20 years included in the original analysis) to identify the necessary timespan of opportunistic data required to elucidate long-term trends.