Biological Conservation 142 (2009) 2282–2292

Contents lists available at ScienceDirect

Biological Conservation

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

eBird: A citizen-based bird observation network in the biological sciences

Brian L. Sullivan *, Christopher L. Wood, Marshall J. Iliff, Rick E. Bonney, Daniel Fink, Steve Kelling

Cornell Lab of Ornithology, 159 Sapsucker Woods Road, Ithaca, NY 14850, USA

ARTICLE INFO

Article history:
Received 4 December 2008
Received in revised form 1 May 2009
Accepted 5 May 2009
Available online 28 May 2009

Keywords:
eBird
Citizen-science
Observation network
Scale-dependent analysis
Bird observations

ABSTRACT

New technologies are rapidly changing the way we collect, archive, analyze, and share scientific data. For example, over the next several years it is estimated that more than one billion autonomous sensors will be deployed over large spatial and temporal scales, and will gather vast quantities of data. Networks of human observers play a major role in gathering scientific data, and whether in astronomy, meteorology, or observations of nature, they continue to contribute significantly. In this paper we present an innovative use of the Internet and information technologies that better enhances the opportunity for citizens to contribute their observations to science and the conservation of bird populations. eBird is building a web-enabled community of bird watchers who collect, manage, and store their observations in a globally accessible unified database. Through its development as a tool that addresses the needs of the birding community, eBird sustains and grows participation. Birders, scientists, and conservationists are using eBird data worldwide to better understand avian biological patterns and the environmental and anthropogenic factors that influence them. Developing and shaping this network over time, eBird has created a near real-time avian data resource producing millions of observations per year.

1. Introduction

Achieving insight about ecological patterns often requires the study of natural systems at large scales (Wiens, 2005). While many studies have been compromised by this challenge (Brown, 1995), an emerging cyberinfrastructure for data synthesis and analysis now permits the collection and organization of data across continent-scale networks that can be used for scale-dependent analysis (Michener et al., 2001). The United States National Science Foundation (NSF) has recognized the significance of scale-dependent ecological analysis and supports continental data-collection efforts through various programs such as the National Ecological Observatory Network (NEON) (http://www.neoninc.org/) and the Long Term Ecological Research network (LTER) (http://www.lternet.edu/). Many data-collection efforts, such as those being developed by the NEON initiative, are autonomous sensor networks. Because most such networks cannot yet identify organisms to species, they serve to gather information on the variables that influence species occurrence (Hochachka et al., 2007). Most data on species-level occurrence still must be gathered by humans (Kelling, 2008), necessitating innovative programs for wide-scale data collection and analysis.

To address this need, many citizen-based observation networks are being developed to gather information on a diverse array of taxa and natural processes. For example, the US National Phenology Network (http://www.usanpn.org/) runs Project Budburst, a citizen-based effort whereby observers report phenological events such as first leafing, first flowering, and first fruit ripening for a variety of plant taxa in order to better understand the broad scale effects of climate change. The Galaxy Zoo (http://www.galaxyzoo.org/) provides access to almost 250,000 images of galaxies and engages volunteers to classify them into shapes in order to better understand how galaxies are formed. The Reef Environmental Education Foundation (REEF; http://www.reef.org/) runs citizen-science projects such as the Great Annual Fish Count to monitor fish populations using amateur divers as sensors.

No organism lends itself more readily to the concept of citizen participation in data gathering than birds. This is because there are nearly 10,000 species that occupy all terrestrial and most aquatic environments and because birds are linked to biotic processes at many levels. Birds are largely diurnal, behaviorally and morphologically conspicuous, and plentiful; they are easily observed, counted, and are among the most studied of all widespread animal groups (Gill, 2006). They engage in the most spectacular long-distance migrations of any organism on the planet, and in so doing demonstrate the biological integration of seemingly disparate ecosystems around the globe (Able, 1999). But most importantly, birds are sensitive environmental indicators, often heralding key changes in environmental processes or ecosystem health.

Birds are frequently encountered and enjoyed by everyday citizens, and for many of them ‘birding’ is a passion. Amateur ornithologists have long studied birds (Barrow, 1998), and perhaps in no other scientific discipline have amateurs had such an historic

---

* Corresponding author. Tel.: +1 609 694 3280.
E-mail address: bls42@cornell.edu (B.L. Sullivan).

0006-3207/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.biocon.2009.05.006

impact. Recognizing this, several leading bird conservation and research organizations such as the Cornell Lab of Ornithology, the British Trust for Ornithology, National Audubon Society, the Patuxent Wildlife Research Lab of the US Geological Survey, and the Royal Society for the Protection of Birds have identified the value of, and have developed the methodologies for harnessing citizen participation. For example, pioneering citizen-science efforts such as the Christmas Bird Count (National Audubon Society, 2002), the United Kingdom's Breeding Bird Survey (Risely et al., 2008), and the United States Breeding Bird Survey (Sauer et al., 1997) annually engage tens of thousands of participants and continue to reveal strong patterns in long-term bird population trends.

Recent initiatives have made use of the Internet as a tool for efficiently gathering, archiving, and distributing bird information to a wide audience. The Internet has broadened our capacity for community outreach, and has made real-time information exchange possible. Indeed a variety of schemes now exist ranging from country-based efforts (e.g., Denmark, http://www.dofbasen.dk), to region-specific applications (e.g., BirdTrack, http://www.bto.org/birdtrack/), to global efforts to organize all projects into a single data-sharing system (e.g., WorldBirds, http://www.worldbirds.org/). Baillie et al. (2006) have shown that projects with an effort-based data gathering model can be useful for determining migration phenology at large scales. The degree to which each system pursues an effort-based approach varies, and some are better than others at gathering data for analysis, but all at minimum record basic information about birds in space and time. With new systems appearing everyday, it is important to underscore the importance of developing projects with a science-based approach to data gathering.

One such effort is eBird (http://www.ebird.org), a program launched by the Cornell Lab of Ornithology (CLO) and the National Audubon Society in 2002, which engages a vast network of human observers (citizen-scientists) to report bird observations using standardized protocols. eBird's mission is to harness the power of everyday birders in an effort to better understand bird distribution and abundance across large spatio-temporal scales and to identify the factors that influence bird distribution patterns.

eBird is built around the simple concept that each time a birdwatcher raises binoculars, he or she has the opportunity to collect useful data. As a tool, eBird serves both the scientific and birding communities by gathering, organizing, and disseminating observations of birds. These data provide information on species occurrence, migration timing, and relative abundance at a variety of spatial and temporal scales. Also, through the process of informal science education, eBird users become better scientists by understanding and using standardized data-gathering techniques, exploring bird data through visualization tools, and interacting with experts. To effectively engage birders, eBird provides a permanent repository for their observations and a method for keeping track of each user's personal observations, birding effort, and various bird lists.

At its most basic level eBird is a tool for birders. But at its highest level, eBird is a tool for science and conservation. eBird data are available in real-time, creating new opportunities for rapid integration of bird data with other kinds of information. As a conservation tool eBird is being used to monitor birds at the site level (e.g., Important Bird Areas (IBAs) (http://www.audubon.org/bird/iba/index.html) by providing data on bird distribution, seasonal occurrence, and relative abundance. eBird data also can be used to test and enhance species distributional models needed to prioritize areas for conservation actions and to direct species-specific management. eBird's broad spatial and temporal component complements more rigorous ornithological research and monitoring programs, allowing scientists to generate new hypotheses and direct future research efforts based on large amounts of data. Importantly, all eBird data become scientific knowledge by joining networked data available through larger global biodiversity initiatives such as the Avian Knowledge Network (AKN) (Avian Knowledge Network, 2008), Global Biodiversity Information Facility (GBIF) (GBIF Data Portal, 2008), and ORNIS (ORNIS Data Portal, 2008).

Since its release, over 500,000 users have visited eBird. The project has gathered over 21 million bird records submitted on more than 1.6 million checklists. Over 35,000 unique users have entered data into eBird, from more than 180,000 locations across the Western Hemisphere and New Zealand. Because eBird accepts "historic" data, 13% of checklists predate the launch of eBird in November, 2002. Since that time, participation has steadily grown (Fig. 1). The majority of the funding to develop eBird was provided by a National Science Foundation award (NSF ESI-0087760). While this initial development of the eBird cyberinfrastructure was a significant investment, the cost per observation is quite low. For example, in 2008 eBird gathered almost 10 million observations of birds. When considering the annual budget for eBird, the cost per observation is three cents. This cost continues to drop as the number of eBird participants increases.

2. Project design

eBird gathers data on bird occurrence and relative abundance at specific locations via a website available in English, Spanish, and

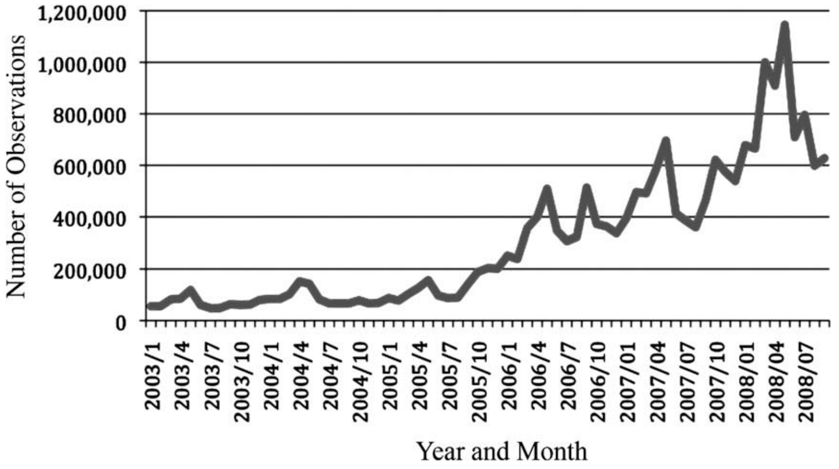

Fig. 1. Number of eBird observations submitted per month since January, 2003. Over 21,000,0000 observations have been gathered to date. Note the peak birding activity in April and May of each year, subsequent annual summer doldrums, and revived interest during fall migration.

French. Users who wish to report bird sightings either choose a location from a drop-down menu of birding "hotspots" (shared locations) or use eBird's online mapping software to select from, or create new, reporting locations. For instance, many participants pinpoint their home as a private location and report birds daily, whereas others bird a local park every day. Chosen locations are stored in the database so that participants can make repeated observations from the same location.

Next, users indicate which of four different protocols they followed while counting birds. Three are effort-based sampling protocols – traveling count, stationary count, and area count – which require associated data such as the amount of time spent birding and distance traveled. The fourth protocol is a less rigorous option, called "Casual observation," which requires only date, location, and species observed to describe the sampling event.

After a user has selected a location and a protocol, eBird then displays a checklist of the species most likely to be observed at the reporting location on the selected datae. The participant then provides the number of individuals seen of each species and submits the completed checklist to the eBird database. The checklist then passes through data quality filters and any unusual records, either birds that are outside of their normal range and/or season, or high numbers of individuals, are flagged for further review. Feedback on flagged records is provided immediately to the submitter (see data verification below) to ensure that information was entered correctly. If the submitter believes their flagged record to be accurate it is then passed on to a regional expert for review and acceptance into the database. Once data have passed through eBird's rigorous filters they can be viewed and summarized by anyone with access to the Internet. Summary tools available in eBird synthesize data to provide useful output for birders. Finally, all eBird raw data are made available via the AKN, where users can choose from prepackaged options or write their own queries directly to the database.

Some participants report birds only occasionally, whereas others submit complete checklists of all the species they see each day, often from several locations – home, office, vacation spot, or favorite birding location. Together these data provide new information on the distribution and abundance of bird populations at both the backyard and continental levels.

While the birding community has traditionally been driven by the search for rarities, eBird also encourages birders to report common birds. eBird therefore provides the basis for gaining a better understanding of the status and distribution of species both common and rare. Moreover, eBird encourages significant detail on unusual reports, which traditionally have been single bird records accompanied by little or no information about the effort involved in obtaining the sample.

Finally, eBird gathers "absence" data by asking participants whether they are reporting all the species that they saw or heard on each checklist. When participants confirm that they are reporting all the species that they observed, we also know which species have not been detected. While these data clearly provide information about true species presence, the challenge is to use these data to infer species absence. The problem is that non-detections may arise when the species is absent, as well as when an observer fails to detect a species that is actually present. In order to distinguish between these confounding signals, we analyze these data in conjunction with additional information about the observations, to create models that account for variation in the detection process (Link and Sauer, 1999; Caruana et al., 2006). For example, our analyses of data with inferred zeros routinely show (e.g., Caruana et al., 2006; Hochachka et al., 2007; Fink and Hochachka, 2009) that more count effort (e.g., time, distance traveled) leads to higher probabilities of species detection. In summary, while the data collected by eBird do not allow a rigorous estimation of the probability of detection (*sensu* MacKenzie et al., 2006), our experience is that the data do provide valuable information on true species occurrence.

2.1. Geographic scope

Currently eBird covers all of the Western Hemisphere and New Zealand, and ultimately will be available worldwide. To ensure relevance at local scales, eBird is managed by local partners through regional portals (hereafter "portals"). Our partners, for example the Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO) in Mexico, ensure that each eBird portal uses a high level of expertise, promotion, and project ownership to better engage local audiences. Each eBird portal is customizable to address local audience needs through specially tailored application functionality, content, and language preferences (e.g., local common names are available for every species for each country). Each portal is fully integrated within the eBird database and application infrastructure so that data can be shared and analyzed freely across both political and geographic boundaries.

2.2. Data Verification

Proper species identification is crucial in all field observation studies. eBird contains a two-stage verification system: (1) instantaneous automated evaluation of submissions based on species count limits for a given date and location; (2) a growing network of more than 500 regional editors composed of local experts who vet records flagged by the automated filters. While it is essential to let people report the birds they see without a complicated data entry process, it is also critical to verify unusual sightings based on geography, exceptional counts, and in the simplest case, extreme rarity.

eBird's automated filters function behind the scenes at the database level. During the data entry process they provide the user with a checklist of the most probable species expected on the reporting date and location. The filters then check user-entered data against average daily count limits for each species before adding it to the database. The count limits reflect the total number of each species likely to be encountered on an average day's birding in a given region. For example, if a count of 10 Western Wood-Pewees (*Contopus sordidulus*) is an acceptable daily total in San Diego County, California in May, but a higher count would be considered exceptional and require further verification, then the count limit for that location and date would be 10. Any record submitted that exceeds this limit prompts the user for confirmation, and if confirmed is sent into a queue for processing by the appropriate regional editor. A user reporting a record of a new species for a given region can add it to his or her checklist using the "Add a species" field. Any record added to a checklist is automatically flagged for review. Many states have county-based filters and some even have sub-county filters based on unique biogeography or avifauna (e.g., Farallon Islands, California). The filters are built and maintained by the regional editors who continually add an ever-increasing level of quality to the eBird data verification process by refining filters as new data become available. Of the 21 million observations to date, roughly 1.2 million have been flagged for review by the automated filters. Of these, our regional editors have validated approximately 62%.

2.3. Obtaining large quantities of data

In ecology and conservation biology, citizen-science techniques provide the opportunity to enlist the public to help survey entire landscapes over long periods (Bhattacharjee, 2005). Citizen-science engages a diversity of participants that range from trained observers to interested citizens, who currently gather tens of millions of observations annually (Bonney, 2007; Kelling, 2008).

Citizen-science projects are typically founded on the thinking that participants will willingly donate their resources, both time and money, to a project whose sole reward is the self-satisfaction inherent in doing something to benefit science. Project designs are based on the formation of a scientific question, and then participants collect data to help find the answer. But too often projects are developed with little or no consideration given to the things that interest participants, and ultimately little or no user incentive or reward is built into the process. Many projects struggle to engage participants, and it can be especially hard to sustain participation. Often when projects incorporate little or no user-reward, the number of participants generally plateaus once the threshold of like-minded individuals is reached. eBird has followed a different model, focused mainly on providing services that appeal to project participants and the birding community, in the process ingesting vast amounts of data. Many projects have asked the question, "what can birders do for science?", but none have asked, "how can we build a useful resource for birders while also engaging them in science?" The shift to the latter model has resulted in expansive eBird growth, both in terms of the number of participants and the amount of data submitted. eBird participation has increased markedly since fall 2005 (Fig. 1) when this paradigm shift occurred. In September of that year we launched upgrades specifically designed to improve user-reward. The result has been that eBird now contributes more data to biodiversity access and analysis initiatives (such as GBIF) than any single project in existence worldwide (GBIF data portal, 2008).

The birding community is driven by the desire to find and identify birds, as well as the recognition and accolades that occur as a result of their discoveries. It is rife with healthy competition, pushing birders to the far ends of the earth in order to find and identify birds. With eBird, we have tapped into this self-motivation by building tools that provide user-reward, ultimately creating, sustaining, and growing participation. An example of this are the "listing" features found in eBird. eBird assigns a suite of geographic values to each checklist submitted, thereby creating user-specific bird lists that are viewable across geographic regions and periods. This simple tool taps into the birding community's basic need to keep and manage bird lists, but also creates incentive for birders to enter both current and historic data by providing exposure and recognition for their work.

Tapping into the competitive side of birding has also increased participation. We have built tools that showcase an individual's data and provide recognition for individual efforts, thereby enhancing personal reward and recognition throughout the birding community. Output tools highlighting the first time that a species has been reported in a geographic region along with the name of the observer who found it build respect and credibility, and create an increased sense of community among eBird users.

These simple changes to what was once a typically designed citizen-science project have resulted in strong eBird growth over the past three years (Fig. 1). By continuing to develop eBird to serve the birding community, we can gather vast and continually growing spatio-temporal data resources.

3. eBird Data Visualization Tools

eBird contains an array of data visualization and analysis tools that provide birders, land managers, and scientists with summary information about bird distribution and relative abundance. It gathers data that help reveal large-scale biological patterns such as distribution changes, relative abundance, and, over time, population trends. eBird complements more rigorous scientific studies by helping to generate new hypotheses, focusing research questions, and adding to the power of observation-based data models. In its visualizations eBird uses "frequency of detection," i.e., the frequency of submitted checklists that report the species of interest. We believe that for most species this metric provides a conservative estimate of the spatial and temporal patterns of abundance both within and across regions. Specifically, it provides the viewer with an indication of how often a bird was detected, without having to consider the actual number of individuals counted. In addition, eBird provides several options to visualize data. For any date range or region the user can select measures of abundance, the average number of birds reported on all checklists; birds per hour, the average number of birds seen per hour spent birding; average count, the average number of birds seen on checklists with a positive observation for the species; high count, the highest count of a species submitted on a single checklist; and total count, the sum of all observations of a species from all checklists. By providing these different metrics the user can begin to understand the dynamic patterns of species occurrence across space and through time.

3.1. Range maps

Because each eBird observation is recorded at a specific location, eBird can generate maps depicting species distribution at multiple spatio-temporal scales. And because eBird gathers large volumes of data, its mapping visualizations are spectacular at both the continental (Fig. 2) and local levels. A 100 km grid layer shows frequency of detection in color-coded squares on the North American map. When visualizing data at the local level simple presence/absence is provided. eBird maps are further enhanced by their inclusion of effort data, which means that eBird shows not only where birds were detected (positive observations for a species) but also where they were not detected (from checklists that do not report a given species but do indicate that they include all species observed in that location at that time).

3.2. Temporal Distribution Patterns

To help users visualize temporal distribution patterns in a familiar way, eBird provides "bar charts" (i.e., frequency histograms) based on frequency of detection for individual species.

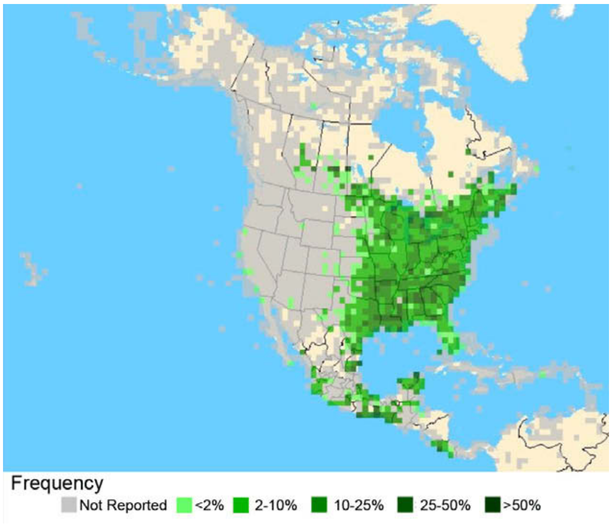

Fig. 2. Frequency distribution of the Ruby-throated Hummingbird (*Archilochus colubris*) across North America. Note the largely eastern distribution in the United States and Canada, and the primary winter grounds from western Mexico south through Costa Rica.

The charts are available at numerous spatial and temporal scales, chosen by the user, from local sites such as Important Bird Areas or nature trails to entire Bird Conservation Regions (Fig. 3.). These visualizations provide users with occurrence information at specific locations at 1-week increments and indicate the likelihood of detecting a species based on its frequency in that area (darker and wider bars indicate increased frequency).

3.3. Regional Statistics

eBird creates a variety of tabular region-based statistics that showcase user effort and create user-reward. “High Counts,” ”All-time firsts and lasts,” and ”Arrivals and Departures” help birders explore the database for records of note. These tools are designed to provide the kinds of output that birders find interesting on a regional basis. For example, one can explore all the record high counts for a US County and see who made each observation on what date and at what location. Because user names are attached to significant records, the competitive nature of birders inspires further data collection. Driven by the peer-recognition of their efforts, birders are encouraged to enter data both new and old, thereby building increased historic perspective in the database.

4. eBird data use

eBird data are valuable to recreational birders and scientists alike. For birders, eBird provides easy access to information about birds in real-time. For scientists, eBird provides valuable bird occurrence information for more than 180,000 locations across the landscape in an organized and accessible format. Below we provide examples of how eBird data and the above-described data visualization tools can be used to discover patterns in bird distribution and abundance at varied spatio-temporal scales, and consider possible conservation applications.

4.1. Visualizing seasonal distribution changes

eBird provides a valuable resource for exploring the basic seasonal distribution of North American birds, many of whose distribution patterns are still poorly known. Because data are gathered year-round, temporally refined maps show breeding, migration, and wintering ranges. For example, an aggregation of frequency maps showing seasonal distribution of Nashville Warbler (*Vermivora ruficapilla*) across North and Central America is presented in Fig. 4. Note the high latitude breeding distribution in the East and the disjunct breeding population, representing the subspecies *V. r. ridgwayi*, in the West. This data visualization also reveals a neatly defined spring migration route through coastal Texas and then north across the Midwest and Appalachians. Fall migration shows a similar distribution pattern, with few records from the Southeastern United States and a small number of birds likely over-wintering in Florida. The species primarily winters in Mexico.

4.2. Monitoring avian range changes

Bird distribution and abundance are constantly changing in response to environmental and anthropogenic factors. eBird provides an excellent platform for studying such change. Consider the Eurasian Collared-Dove (*Streptopelia decaocto*). Accidentally introduced from Eurasia to the Bahamas in the late 1970s, this species has now overtaken much of North America (Romagosa and McEneaney, 2000; Romagosa, 2002). Using eBird, we can visualize the expansion using frequency maps, which are updated with new records every 24 h (Figs. 5 and 6). The maps show that the species moved quickly from the Bahamas to the western United States, but expanded only partially into the northeastern states. This pattern is similar to that shown by the Eurasian Collared-Dove as it expanded rapidly west–northwest across Europe and then slowly backfilled into Asia (Fisher, 1953). Because eBird collects data on all species, it is effective for monitoring the large-scale distribution patterns of native North American birds as well as distribution patterns of introduced avifauna that might negatively impact native species in coming years.

4.3. Differential migration timing

eBird data can be used to examine the timing of migration across a large geographic area. The Yellow Warbler (*Dendroica petechia*) is a widespread breeder across North America and a long-distance Neotropical migrant with at least eight migratory subspecies breeding in or passing through the United States (Lowther et al., 1999). To determine if these populations show differences in migration timing across similar latitudes, we used eBird to compare the average peak migration periods for Yellow Warblers at 39° latitude at three locations (Fig. 7). Spring arrival times differed by roughly one week, peaking earliest in California (8 May in Sacramento), then New Jersey (15 May in Cape May), and finally along the Colorado front range (22 May). In the fall, peak migration starts on the early date of 1 August in Cape May County, where the majority of migrants are the proximal *D. p. aestiva*, but is much later in California, peaking on 15 September in Sacramento, where

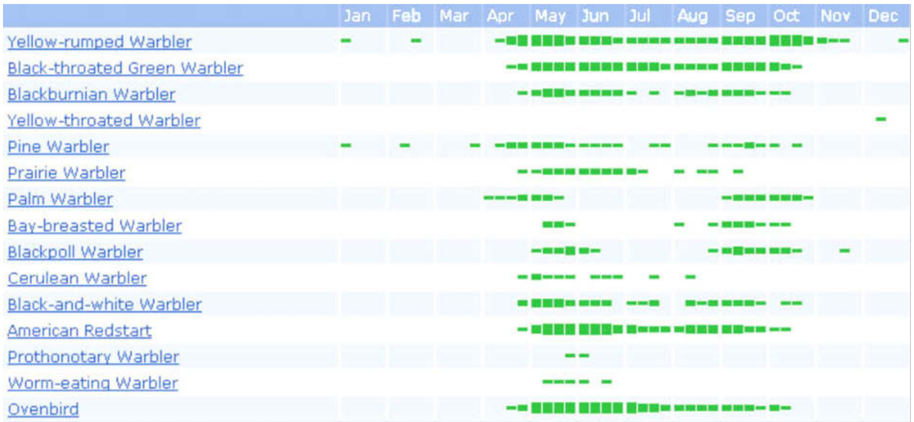

Fig. 3. eBird frequency chart showing the seasonal distribution of a subset of Wood-Warblers for Tompkins County, NY, USA. In this example many species show well-defined migration intervals, with most peaking in mid-May.

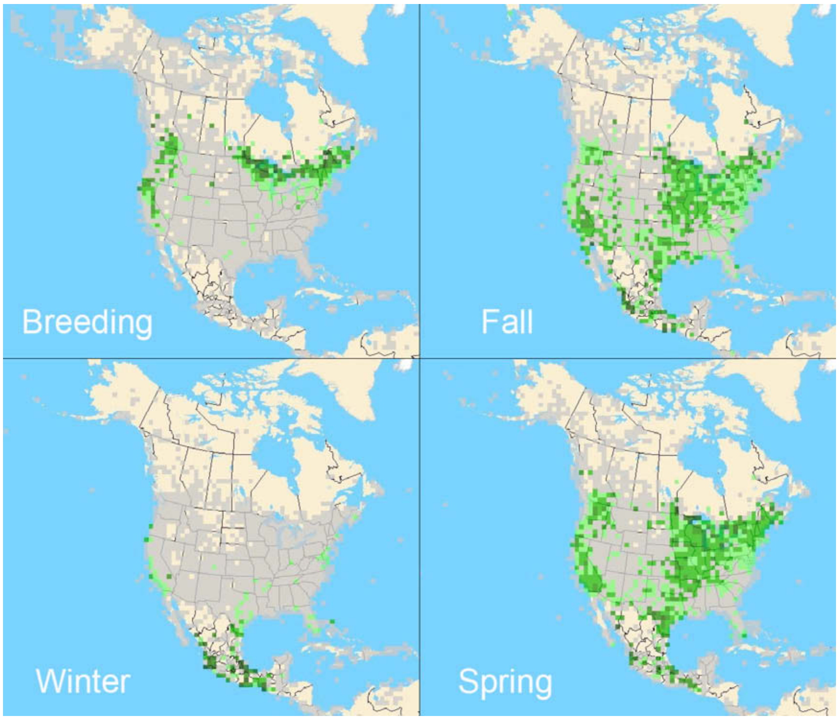

Fig. 4. Seasonal changes in Nashville Warbler (*Vermivora ruficapilla*) distribution in North America aggregated across years. When the user requested these maps, eBird created them by dividing the continent into 100 km × 100 km blocks. If at least five checklists were submitted from a given block for the particular season then the block was filled, either with green to indicate presence or gray to indicate that checklists were submitted but no Nashville Warblers were reported. Darker shades of green indicate higher frequency of checklists reporting Nashville Warbler within a block. Thus the maps provide information both on coverage (e.g., where data were collected) as well as an indication of how common the species was. eBird used a total of 1,089,676 checklists to create this visualization, and 24,596 of the blocks recorded Nashville Warbler. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

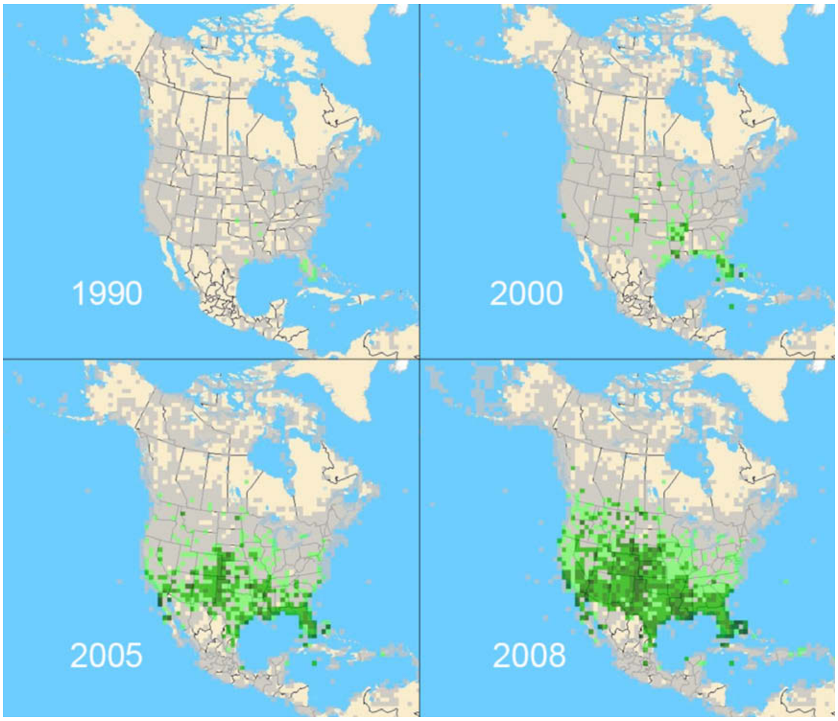

Fig. 5. Eurasian Collared-Dove (*Streptopelia decaocto*) frequency distribution across North America from 1990 to 2008. In this spatio-temporal view of the species' expansion, note the massive overall expansion from its origin in Florida across northern and western North America, and its surprisingly limited expansion into the northeastern United States.

late-breeding individuals represent more northerly populations, *D. p. rubiginosa*, breeding in Alaska (Grinnell and Miller, 1944). Explorations of these data reveal interesting biological patterns both within and across taxa, stimulate hypothesis generation, and

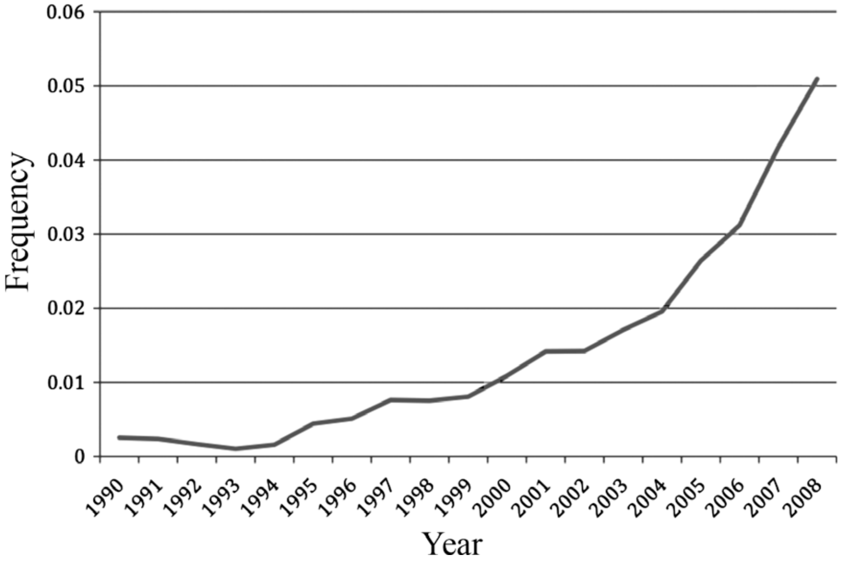

Fig. 6. Eurasian Collared-Dove (*Streptopelia decaocto*) frequency across the North America states from 1990 to 2008.

inspire critical thinking. Over time, such explorations are expected to reveal unknown biological phenomena in species for which there is currently little information.

4.4. eBird as a tool for priority species conservation

eBird has a broad user-base spanning multiple continents. As such, it can be a powerful tool when put to use for targeted data gathering efforts. Rusty Blackbirds (*Euphagus carolinus*) are in serious decline, by some estimates as much as 70–90% over the last 40 years (Greenberg and Droege, 1999; Niven et al., 2004; Sauer et al., 2008). Scientists have cited a need for an increased understanding of this species' natural history, especially during migration and winter. Recognizing the broad capacity of eBird to gather data, the Rusty Blackbird Working Group created the Rusty Blackbird Blitz, a citizen-science effort whose data gathering capacity is founded in eBird. Over the course of a 9-day period in early February, thousands of eBird users attempt to find Rusty Blackbirds just prior to spring migration. These observations are reported to eBird and the data are then disseminated to the working group scientists for analysis. Observations yield information on habitat use, flock size, age, sex, and species association. In time this targeted winter survey effort, when combined with migrant Rusty Blackbird observations, will help paint a clearer picture of this species' ecological requirements, and enable conservationists to help preserve it and the habitat it needs.

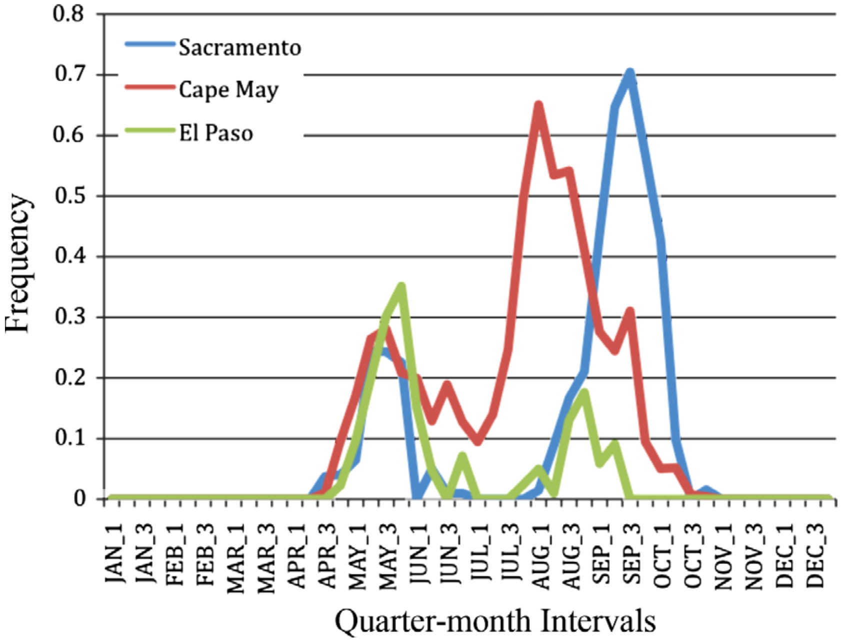

Fig. 7. Seasonal distribution of Yellow Warbler (*Dendroica petechia*) at Sacramento County, California ($N = 2522$ checklists), El Paso County, Colorado ($N = 987$ checklists), and Cape May County, New Jersey ($N = 5071$ checklists), USA. The centroid of each county lies at roughly 39° latitude. This graph uses eBird data to show average differences in arrivals and departures at the three locations. Time periods roughly correspond to weeks, but all months are divided into four week intervals. *Y*-axis shows frequency of checklists reporting Yellow Warbler across all years.

4.5. Delineating migration timing for conservation management

To manage landscapes for optimal bird conservation, land managers need a site-level understanding of bird distribution and abundance. Moreover, temporal migration patterns and site usage by threatened and endangered species and species of conservation concern must be delineated and understood. As an example, managing wetlands for both shorebirds and waterfowl presents an age-old dilemma. Land managers trying to optimize water levels typically have lacked important information such as migration timing for individual species. Using eBird, managers can view aggregated results for their sites (Fig. 8), and quickly understand when species of concern are using the area. The biological patterns revealed could be compared with surrounding sites where birds occur in unaltered landscapes for unbiased reference. Simple frequency of occurrence throughout the year gives a clear view of when shorebird migration peaks in spring and fall and when the bulk of waterfowl arrive. The two peak times are often weeks or even months apart, permitting site management for both groups.

4.6. Providing data resources for decision support tools

A significant strength of eBird as a conservation tool lies in its near real-time collection and dissemination of data. Up-to-the-minute bird observations can be coupled with disparate data sources to provide powerful on-the-ground conservation tools. An example of this is a visualization tool built by PRBO Conservation Science (http://www.prbo.org) that details the distribution of recent bird observations in relation to oiled coastline during the Cosco-Busan oil spill that occurred in San Francisco Bay, California, on 7 November, 2007. The spill impacted at least four Important Bird Areas and rapidly spread to beaches around the mouth of San Francisco Bay. A map-based tool was quickly created depicting an estimate of oiled shoreline based on data collected from the Office of Spill Prevention and Response, coupled with Important Bird Area polygons, and recent bird sightings data collected by eBird. The resulting tool allowed managers to prioritize areas for clean-up that showed high degrees of overlap between known oil distribution and high bird diversity and abundance. The resulting visualization was used to predict which species would be most impacted and where responders should concentrate their efforts to gather birds for treatment.

4.7. Modeling relative abundance

Like many ecological data, eBird counts arise as the product of two distinct, linked processes: detection, or observation, processes and the ecological processes governing abundance. The scientific goal is to control for as much of the detection process as possible in order to reveal the true abundance of the species. Here we provide a brief illustration to show how model-based analysis of eBird data can be used to control for important sources of bias. In this example we study the expected breeding season eBird counts for Northern Cardinal (*Cardinalis cardinalis*), a common bird in eastern North America. eBird traveling counts recorded between 2004 and 2007 with transect distances less than 8.1 km and total on-effort times less than 3 h were modeled using Bagged Decision Trees, a nonparametric model (Hochachka et al., 2007; Fink and Hochachka, 2009).

One important source of bias in eBird data is the highly variable sampling effort expended by participants. By including appropriate covariates, analytic models can provide simultaneous control for several aspects of this bias. Variation in detection rates is modeled as a function of effort spent watching birds, both the total time spent watching birds and the length of the traveling count. The partial dependence plots (Hastie et al., 2001) (Fig. 9) show the estimated effect of total time (hours) and distance (kilometers), after controlling for all other effects in the model. As participants spend more time looking for birds, they are more likely to record higher counts of Northern Cardinal, with diminishing returns as the duration of the search increases. Similarly, once participants reach transects greater than a half kilometer in length, they tend to see more Northern Cardinals the farther they travel. Variation in availability

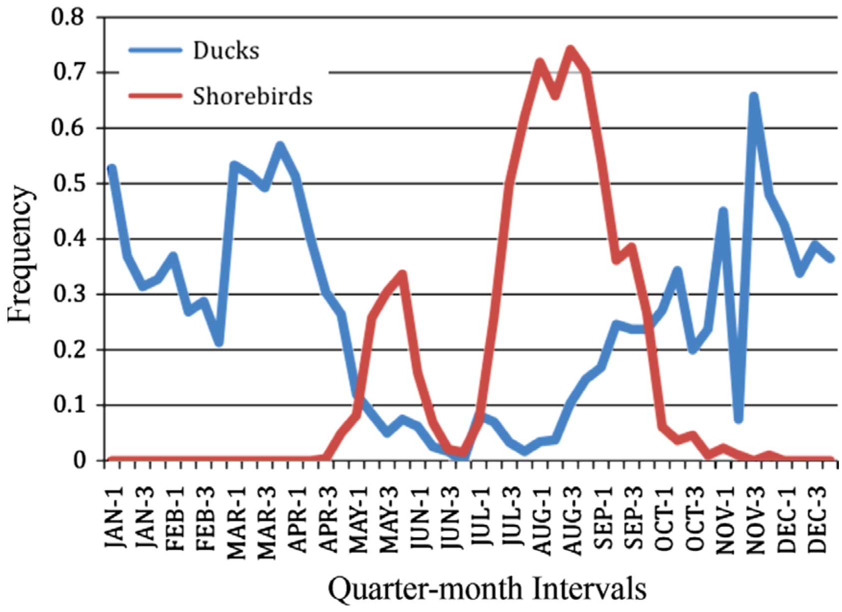

Fig. 8. Seasonal distribution of a subset of five migratory shorebirds (red) and five migratory ducks (blue) at Jamaica Bay IBA in New York, USA ($N = 1064$ checklists). Land managers wishing to manage water levels for target species can easily visualize peak usage and adjust management practices accordingly. For example, lower water levels could be maintained for shorebirds during their peak migration period in May and again from July to September, whereas higher water levels would most benefit ducks during winter and during peak migration periods from late February to April and from late September to December. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

for detection is modeled as a function of the observation time of day and date.

Fig. 10 shows the estimated abundance surface of Northern Cardinal for spring 2006. This surface estimate controls for variation in detection rate by holding observer effort, distance traveled, and time of day constant. The prediction scale is the expected number of Northern Cardinals observed during eBird counts initiated at 6 AM along a 1 km transect conducted over 1 h. This surface confirms the spatial accuracy of the model at the national scale. The surface closely matches the known distribution of Northern Cardinal, showing the highest predicted numbers in the southeastern United States, and the lowest predicted numbers near the northern periphery of its range in New England. The small Arizona population is also modeled. The value of this analysis is its ability to predict relative abundance for a species, and to extend those predictions into areas where we have little or no data. In the future, we hope to refine and strengthen the model to allow predictions for rare species with sparse data, creating new surfaces of relative abundance for threatened and endangered species. The ability to model bird distribution and abundance in areas with sparse data will provide a better understanding of how birds occur in remote or poorly sampled regions.

5. eBird potential biases

While eBird provides important information on species distribution, frequency and relative abundance as shown in the examples above, there are certain biases associated with the data. First, species detectability is a problem in most bird-sampling techniques, and easily detected birds are reported more frequently than cryptic species. While eBird users are asked whether they are submitting a complete checklist of birds for a given date and location, we cannot know for sure whether they detected all the species that were present. Second, eBird appeals to a wide variety of users with varying skill levels, and the database is vulnerable to misidentifications. While all eBird checklists are processed through the same data verification schemes (see above), it would be impossible to consider that all observations are correctly reported. Third, geographic biases exist due to unstratified sampling techniques; eBird allows birders to report their observations from wherever they make them. "Locations" are not reported in any standardized fashion for traveling counts, and while observers are encouraged to plot their point at the middle of the transect, many transects are plotted at the beginning or end of the transect. This makes it difficult to link the birds they've reported with specific habitats on the ground. Finally, the birding community is not evenly distributed across the landscape, and there is bias associated with the distribution of birding effort (Ferrer et al., 2006). The eBird dataset is most heavily concentrated in areas with high human populations and is less extensive in sparsely populated regions.

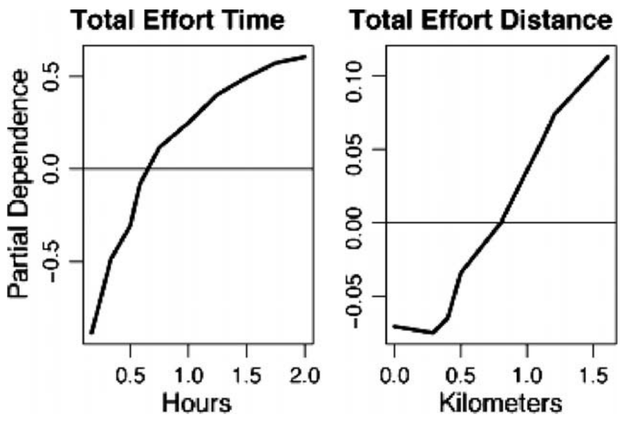

Fig. 9. Partial dependence of total effort time spent on counts that recorded Northern Cardinal (left) and the length of the traveling count (right). Partial effects are used to visually examine the dependence of predictive models on small subsets of predictors, accounting for the averaged effects of the other variables in the model. The partial effect for effort time shows that participants who spend more time looking for birds are likely to record higher counts of Northern Cardinal. Moreover, the effect of each unit of additional time has diminishing returns as the duration of the search increases. Similarly, once participants reach transects a half mile in length, they tend to see more Northern Cardinals the farther they travel.

6. Discussion

eBird provides an example of how a global network of avian biological sensors (i.e., birdwatchers) has been created, maintained, and nurtured, providing millions of observations for analysis, ultimately moving us closer to understanding avian population dynamics in real-time. By providing participants with desired tools and ample reward for their involvement, we are creating and self-sustaining a flow of avian biological data from around the planet. Unlike perhaps any other recreational/outdoor activity, birders are an increasingly powerful and well-trained observational network. Estimates of the total number of birders in North America range from hundreds of thousands to 70 million (La Rouche, 2003; Rich et al., 2005; Leonard, 2008), making birdwatching one of America's favorite recreational pastimes. The amount of information available to birders on the topic of bird identification is astounding, and more information is continually available through new guides and in Internet forums. Birders represent a growing network of skilled observers, and as such we believe that they represent a valuable data collection tool, one that can be shaped into an even better and more science-minded bird-recording force as time goes on.

eBird has potential to help birders become better scientists in many ways. The most obvious education occurs when a novice birder reports something unusual and is questioned by a peer or mentor, thereby learning more about bird identification, distribution, or migration timing in the process. Because eBird has been developed to steer birders toward providing more useful data, a secondary learning process happens in the shift from making 'casual observations' to any of the 'effort-based' methodologies. Our results show that over time eBird users have shifted away from making simple casual observations, and have begun to provide us with more useful effort-based data (Fig. 11). This trend shows that you can teach 'old birders new tricks', and is significant in showing that birders not only want to participate in citizen-science, but are willing to change the way they go birding to provide more robust scientific data.

The true power of eBird lies in the strength and diversity of its users. Everyone with an interest in birds can participate, from the rank novice, to the backyard birder, to the globe-trotting expert. We must strive to encourage more birders to use the application. As more people and more diverse users submit observations to eBird, the utility of its database will vastly improve. Moreover, as more users submit data an environment of sharing and free data exchange will become the norm between birders, scientists, and conservationists. An urgent need is for birders to begin to explore new and uncharted birding territory. By visiting new habitats that are little traveled by birders we can create better species–habitat relationships. More checklists from more locations will allow us to better understand birds, the habitats they require and how to protect them. Beyond the borders of North America, eBird needs to be promoted as a tool for birders, science, and conservation across the larger landscape.

The ability to create, manage, and manipulate vast, real-time data resources is influencing the ways in which we study biology and conduct conservation research and planning. An emerging

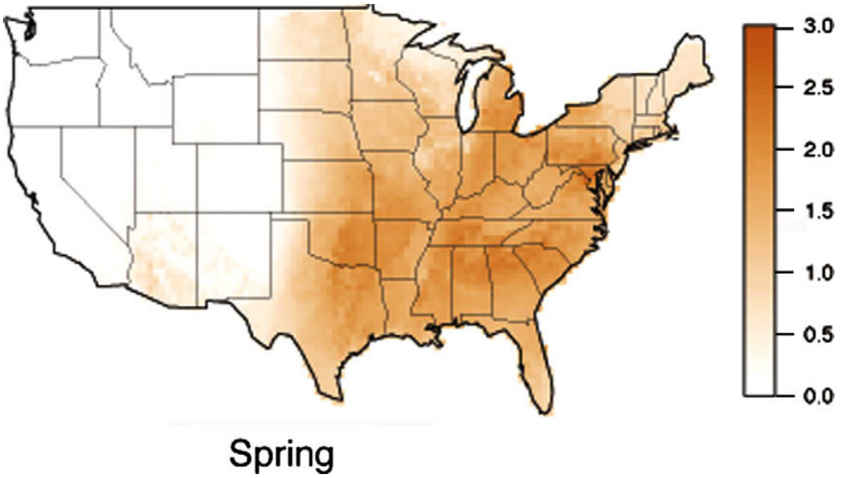

Fig. 10. Northern Cardinal spring 2006 relative abundance map. This surface shows the estimated abundance surface of Northern Cardinal for spring 2006 taking into account local habitat characteristics as well as controlling variation in detection rate by holding observer effort, distance traveled, and time of day constant. The prediction scale is the expected number of Northern Cardinals per eBird count observed at 6 AM on a 1 km transect conducted over 1 h.

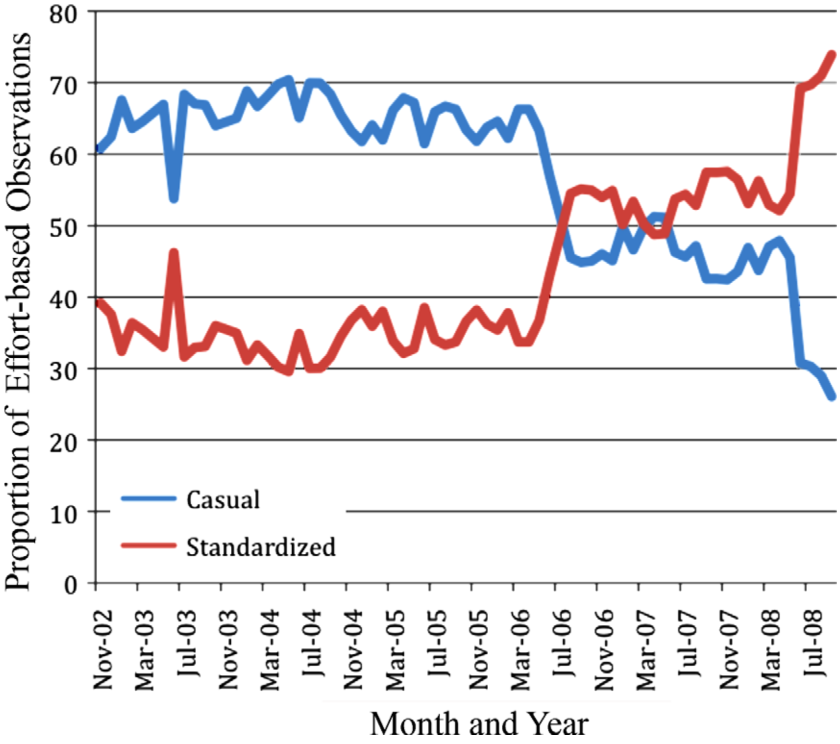

Fig. 11. Proportion of checklists submitted by protocol from November, 2002 to September, 2008. This graph shows the shift from a relatively high proportion of "Casual observations" (i.e., observations without effort) during the first three years of eBird data collection, to more widespread use of the three 'effort-based' methodologies grouped together here under the term "Standardized." The two precipitous changes reflect separate targeted outreach efforts through which users were educated about the importance of submitting standardized observations. The significance of this graph is that over time, eBird users have learned that effort-based methodologies are more useful to science, and have changed the way they go birding to accommodate effort-based requirements.

cyberinfrastructure consisting of databases, network protocols, and computational services (Stein, 2008) is changing the way we collect, store, and analyze biological data. To take full advantage of the broad spatio-temporal possibilities offered to conservationists in the "Big Data Age" (Nelson, 2008), a global network of biological "sensors" must be built containing both amateur and expert observers. eBird is a excellent example of how new technologies can be used to grow, manage, and sustain such a network, whose cooperation may be the only way to gather large amounts of data, which can be used *a posteriori* to assess questions with spatial and temporal coverage impossible to reach by individual initiatives. By collecting a variety of biological data this network will allow us to synthesize massive amounts of information in real-time, providing answers to important and long-standing biological questions (e.g., population trends, distribution, abundance, and demographics). An important next step will be the continual use of the accumulating data to improve our understanding of these biological processes and how they are affected by anthropogenic and environmental factors. Armed with a "Big Picture" perspective of avian population dynamics we can better engage policy makers and achieve desired conservation outcomes by proposing reasonable, understandable, and accountable solutions. By gathering, analyzing, and interpreting massive amounts of data collected by autonomous sensors, trained scientists, and birders, we hope to better understand the conservation issues facing birds and biodiversity across the planet.

Acknowledgements

We'd like to thank all eBird participants, without whom the project would not be possible. We acknowledge the Information Science group at the Cornell Lab of Ornithology for providing support for eBird, especially the following people: T. Lenz, J. Gerbracht, P. Allen, W. Hochachka, T. Fredericks, T. Levatich, R. Slothower, D. Danowski, K. Webb, W. Morris, G. Delisle, J. Fitzpatrick, as well as two anonymous reviewers whose comments greatly improved this manuscript. Finally we wish to thank the hundreds of birders who have lent their regional expertise to our data quality process through creating filters, reviewing records, and critiquing eBird itself. Without their continued support and assistance, and without the participation of tens of thousands of birders, eBird would not exist.

Development of eBird was supported in part by the National Science Foundation under Grant ESI-0087760. Additional support for the analysis of eBird data came from the Leon Levy Foundation, Wolf Creek Foundation, and the National Science Foundation (Grant Numbers ITR-0427914, DBI-0542868, DUE-0734857, and IIS-0612031). Any opinions, findings, and conclusions or recommendations expressed in this manuscript are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Able, K., 1999. Gatherings of Angels: Migrating Birds and Their Ecology. Comstock Press.

Avian Knowledge Network, 2008. Avian Knowledge Network: An Online Database of Bird Distribution and Abundance [web application]. Ithaca, New York. <http://www.avianknowledge.net> (accessed 08.06.08).

Baillie, S.R., Balmer, D.E., Downie, I.S., Wright, K.H.M., 2006. Migration watch: an Internet survey to monitor spring migration in Britain and Ireland. Journal of Ornithology 147, 254–259.

Barrow, M.V., 1998. A Passion for Birds: American Ornithology After Audubon. Princeton University Press, Princeton, New Jersey.

Bhattacharjee, Y., 2005. Citizen scientists supplement work of Cornell researchers. Science 308, 1402–1403.

Bonney, R., 2007. Citizen science at the Cornell Lab of Ornithology. In: Yager, Robert E., Falk, John H. (Eds.), Exemplary Science in Informal Education Settings: Standards-Based Success Stories. NSTA Press. 278 pp (Chapter 16).

Brown, J.H., 1995. Macroecology. University of Chicago Press, Chicago. 248 pp.

Caruana, R., Elhawary, M., Monson, A., Riedewald, M., Sorokina, D., Fink, D., Hochachka, W.M., Kelling, S., 2006. Mining citizen science data to predict abundance of wild bird species. In: Ungar, L., Craven, M., Gunopulos, D., Eliassi-Rad, T. (Eds.), 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM Press, pp. 909–915.

Ferrer, X., Carrascal, L.M., Gordo, O., Pino, J., 2006. Bias in avian sampling effort due to human preferences: an analysis with Catalonian birds (1900–2002). Ardeola 53, 213–227.

Fink, D., Hochachka, W., 2009. Gaussian semiparametric analysis using hierarchical predictive models. Environmental and Ecological Statistics 3, 1011–1035.

Fisher, J., 1953. The collared turtle dove in Europe. British Birds 5, 153–181.

GBIF Data Portal. <www.gbif.net> (accessed 08.06.08).

Gill, F., 2006. Ornithology. W.H. Freeman.

Greenberg, R., Droege, S., 1999. On the decline of the Rusty blackbird and the use of ornithological literature to document long-term population trends. Conservation Biology 13, 553–559.

Grinnell, J., Miller, A.H., 1944. The Distribution of the Birds of California. Artemisia Press, Lee Vining, CA.

Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Verlag, New York.

Hochachka, W., Caruana, R., Fink, D., Munson, A., Riedewald, M., Sorokina, D., Kelling, S., 2007. Data mining for discovery of pattern and process in ecological systems. Journal of Wildlife Management 71, 2427–2437.

Kelling, S., 2008. Significance of organism observations: Data discovery and access in biodiversity research. Report Presented to the Global Biodiversity Information Facility, Copenhagen.

La Rouche, G.P., 2003. Birding in the United States: A Demographic and Economic Analysis. Report 2001-1. US Fish and Wildlife Service. Washington, DC.

Leonard, J., 2008. Wildlife Watching in the US: The Economic Impacts National and State Economies in 2006. US Fish and Wildlife Service, Arlington, Virginia.

Link, W.A., Sauer, J.R., 1999. Controlling for varying effort in count surveys–an analysis of Christmas bird county data. Journal of Agricultural, Biological and Environmental Statistics 4, 116–125.

Lowther, P.E., Celada, C., Klein, N.K., Rimmer, C.C., Spector, D.A., 1999. Yellow Warbler (*Dendroica petechia*). In: Poole, A. (Ed.), The Birds of North America Online. Cornell Lab of Ornithology, Ithaca (Retrieved from the Birds of North America Online. http://bna.birds.cornell.edu/review/species/454).

MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L., Hines, J.E., 2006. Occupancy Estimation and Modeling. Academic Press, Burlington, MA.

Michener, W., Baerwald, T., Firth, P., Palmer, M., Rosenberger, J., Sandlin, E., Zimmerman, H., 2001. Defining and unraveling biocomplexity. Bioscience 51, 5.

National Audubon Society, 2002. The Christmas Bird Count Historical Results. <http://www.audubon.org/bird/cbc> (accessed 08.06.08).

Nelson, S., 2008. Big data: the Harvard computers. Nature 455, 36–37.

Niven, D.K., Sauer, J.R., Butcher, G.S., Link, W.A., 2004. Christmas bird count provides insights into population change in land birds that breed in the boreal forest. American Birds 58, 10–20.

ORNIS data portal. <http://ornisnet.org> (accessed 08.06.08).

Rich, T. D., Beardmore, C.J., Berlanga, H., Blancher, P.J., Bradstreet, M.S.W, Butcher, G.S., Demarest, D.W., Dunn, E.H., Hunter, W.C., Iñigo-Elias, E.E., Kennedy, J.A., Martell, A.M., Panjabi, A.O., Pashley, D.N., Rosenberg, K.V., Rustay, C.M., Wendt, J.S., Will, T.C., 2005. Partners in Flight North American Landbird Conservation Plan. Cornell Lab of Ornithology, Ithaca, NY. Partners in Flight Website. <http://www.partnersinflight.org/cont_plan/> (version: March 2005).

Risely, K., Noble, D.G., Baillie, S.R., 2008. The breeding bird survey 2007. British Trust for Ornithology.

Romagosa, C.M., 2002. Eurasian Collared-Dove (*Streptopelia decaocto*). In: Poole, A. (Ed.), The Birds of North America Online. Cornell Lab of Ornithology, Ithaca (Retrieved from the Birds of North America Online. http://bna.birds.cornell.edu/review/species/630).

Romagosa, C.M., McEneaney, T., 2000. Eurasian Collared-Dove in North America and the Caribbean. North American Birds 53, 348–353.

Sauer, J.R., Hines, J.E., Gough, G., Thomas, I., Peterjohn, B.G., 1997. The North American Breeding Bird Survey Results and Analysis. Version 96.4. Patuxent Wildlife Research Center, Laurel, MD.

Sauer, J.R., Hines, J.E., Fallon, J., 2008. The North American Breeding Bird Survey, Results and Analysis 1966–2007. Version 5.15.2008. USGS Patuxent Wildlife Research Center, Laurel, MD.

Stein, L.D., 2008. Towards a cyberinfrastructure for the biological sciences: progress, visions and challenges. Nature Reviews Genetics 9, 678–688.

Wiens, J., 2005. Toward a unified landscape ecology. In: Wiens, J., Moss, M. (Eds.), Issues and Perspectives in Landscape Ecology. Cambridge University Press, Cambridge, pp. 365–373.