2018

# Opportunities and challenges for big data ornithology

Frank A. La Sorte
*Cornell University*, fal42@cornell.edu

Christopher A. Lepczyk
*Auburn University*

Jessica L. Burnett
*University of Nebraska-Lincoln*, jburnett@huskers.unl.edu

Allen H. Hurlbert
*University of North Carolina at Chapel Hill*

Morgan W. Tingley
*University of Connecticut*

***See next page for additional authors***

**Authors**
Frank A. La Sorte, Christopher A. Lepczyk, Jessica L. Burnett, Allen H. Hurlbert, Morgan W. Tingley, and Benjamin Zuckerberg

THE CONDOR
Ornithological Applications
AmericanOrnithology.org

Volume 120, 2018, pp. 414–426
DOI: 10.1650/CONDOR-17-206.1

*REVIEW*

# Opportunities and challenges for big data ornithology

**Frank A. La Sorte,[1]* Christopher A. Lepczyk,[2] Jessica L. Burnett,[3] Allen H. Hurlbert,[4,5] Morgan W. Tingley,[6] and Benjamin Zuckerberg[7]**

[1] Cornell Lab of Ornithology, Cornell University, Ithaca, New York, USA
[2] School of Forestry and Wildlife Sciences, Auburn University, Auburn, Alabama, USA
[3] Nebraska Cooperative Fish and Wildlife Research Unit, School of Natural Resources, University of Nebraska Lincoln, Lincoln, Nebraska, USA
[4] Department of Biology, University of North Carolina, Chapel Hill, North Carolina, USA
[5] Curriculum for the Environment and Ecology, University of North Carolina, Chapel Hill, North Carolina, USA
[6] Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, Connecticut, USA
[7] Department of Forest and Wildlife Ecology, University of Wisconsin–Madison, Madison, Wisconsin, USA
* Corresponding author: fal42@cornell.edu

Submitted October 10, 2017; Accepted February 22, 2018; Published May 2, 2018

## ABSTRACT

Recent advancements in information technology and data acquisition have created both new research opportunities and new challenges for using big data in ornithology. We provide an overview of the past, present, and future of big data in ornithology, and explore the rewards and risks associated with their application. Structured data resources (e.g., North American Breeding Bird Survey) continue to play an important role in advancing our understanding of bird population ecology, and the recent advent of semistructured (e.g., eBird) and unstructured (e.g., weather surveillance radar) big data resources has promoted the development of new empirical perspectives that are generating novel insights. For example, big data have been used to study and model bird diversity and distributions across space and time, explore the patterns and determinants of broad-scale migration strategies, and examine the dynamics and mechanisms associated with geographic and phenological responses to global change. The application of big data also holds a number of challenges wherein high data volume and dimensionality can result in noise accumulation, spurious correlations, and incidental endogeneity. In total, big data resources continue to add empirical breadth and detail to ornithology, often at very broad spatial extents, but how the challenges underlying this approach can best be mitigated to maximize inferential quality and rigor needs to be carefully considered.

*Keywords:* big data, eBird, citizen science, ornithology, semistructured, unstructured, weather surveillance radar

**Oportunidades y desafíos para la ornitología de los datos masivos**

## RESUMEN

Los avances recientes en la tecnología de la información y la adquisición de datos han creado tanto nuevas oportunidades de investigación como desafíos para el uso de datos masivos (big data) en ornitología. Brindamos una visión general del pasado, presente y futuro de los datos masivos en ornitología y exploramos las recompensas y desafíos asociados a su aplicación. Los recursos de datos estructurados (e.g., Muestreo de Aves Reproductivas de América del Norte) siguen jugando un rol importante en el avance de nuestro entendimiento de la ecología de poblaciones de las aves, y el advenimiento reciente de datos masivos semi-estructurados (e.g., eBird) y desestructurados (e.g., radar de vigilancia climática) han promovido el desarrollo de nuevas perspectivas empíricas que están generando miradas novedosas. Por ejemplo, los datos masivos han sido usados para estudiar y modelar la diversidad y distribución de las aves a través del tiempo y del espacio, explorar los patrones y los determinantes de las estrategias de migración a gran escala, y examinar las dinámicas y los mecanismos asociados con las respuestas geográficas y fenológicas al cambio global. La aplicación de datos masivos también contiene una serie de desafíos donde el gran volumen de datos y la dimensionalidad pueden generar una acumulación de ruido, correlaciones espurias y endogeneidad incidental. En total, los recursos de datos masivos continúan agregando amplitud y detalle empírico a la ornitología, usualmente a escalas espaciales muy amplias, pero necesita considerarse cuidadosamente cómo los desafíos que subyacen este enfoque pueden ser mitigados del mejor modo para maximizar su calidad inferencial y rigor.

*Palabras clave:* ciencia ciudadana, datos masivos, desestructurado, eBird, ornitología, semi-estructurado, vigilancia climática

## Background

Throughout the history of ornithology, dedicated natural historians and professional ornithologists have made use of an ever-expanding set of tools to derive insights into and increase the breadth of our understanding of avian biology. The shotgun gave way to binoculars, then mist nets, leg bands, and geolocators. Likewise, stable isotopes have provided new insights into diets and distributions, and DNA sequencing has improved our understanding of population structure and evolutionary history. Added to these technological and inferential advances, "big data" are poised to vastly improve our understanding of the distributions and ecology of birds.

**Defining big data.** The term "big data" is used to describe the large digital datasets that have emerged as a result of recent advancements in information technology and data acquisition (Tien 2013). Big data resources are often differentiated from traditional data sources based on 3 features: (1) data are numerous; (2) data are generated, captured, and processed rapidly; and (3) data cannot be readily organized into a traditional relational database (Hashem et al. 2015). The concept of big data, however, has advanced beyond identifying large and complex datasets to describing the broader cultural, technological, and scholarly implications of the growth of these unique resources (Boyd and Crawford 2012). The big data phenomenon has permeated many social, political, and commercial domains, as well as many scientific and medical disciplines (Boyd and Crawford 2012), and this progression has resulted in paradigm shifts that are revolutionizing many aspects of human life (Mayer-Schönberger and Cukier 2013).

A common conceptual framework for describing the characteristics of individual big data resources is the 3 V's: volume, velocity, and variety (Laney 2001). Volume refers to the amount of data collected. Velocity refers to the rate at which data are collected. Variety refers to the structural heterogeneity in a dataset, which is typically classified as structured, semistructured, or unstructured (Gandomi and Haider 2015). Structured data, which constitute a very small proportion by volume of existing big data, are data that can be conveniently stored in traditional spreadsheets or relational databases. Unstructured data, currently the dominant format representing ~80% of existing data, refers to information that lacks the structural organization that is required for efficient storage and analysis. Semi-structured data fall along the continuum between these 2 extremes and are generally defined by more flexible structural elements. Beyond volume, velocity, and variety, 2 additional concepts are useful when considering the application of big data resources: veracity and value (Gandomi and Haider 2015). Veracity refers to data precision and uncertainty, and value refers to the information to volume ratio contained in the data. Successfully recognizing and managing these characteristics and challenges is necessary before big data can be leveraged to generate knowledge.

**Big data in the natural sciences.** Big data are often classified into 2 categories, data originating from the physical or natural world obtained through observations or sensors, and data obtained from social or economic activities (Jin et al. 2015). Examples of unstructured big data from the physical or natural world include satellite imagery, seismic imagery, astronomical imagery, atmospheric data, and data from high energy physics. Across the natural sciences, including ornithology, big data have challenged current epistemologies through the creation of data-driven methods, such as machine learning, that rely more on exploring large datasets and less on theory or hypothesis testing (Kitchin 2014). The acquisition of data from the natural world has been a characteristic of human history for hundreds of years. With recent technological advancements, the advent of big data can be seen as a logical continuation of this process (Bowker 2000, Strasser 2012). With the advent of big data, the speed and scope of data acquisition in the natural sciences has accelerated rapidly, creating new opportunities and challenges (Hampton et al. 2013, Peters et al. 2014, Schimel and Keller 2015, Devictor and Bensaude-Vincent 2016, LaDeau et al. 2017).

To extract scientific knowledge about the natural world from big data in a robust and efficient fashion requires new analytical approaches and perspectives. The fields of bioinformatics and ecoinformatics emerged in part to address how to extract and apply these types of data to the natural sciences (Sarkar 2009, Michener and Jones 2012). Within these approaches, data can originate from remote-sensing platforms (Jensen 2006) or sensor networks that are embedded in the environment (Porter et al. 2009, Benson et al. 2010). The latter include inanimate sensor networks, such as the U.S. National Ecological Observatory Network (NEON; Keller et al. 2008), and human sensor networks, in which data are collected by the general public (Bonney et al. 2009, Dickinson et al. 2010). Data compiled through these methods are unique in the natural sciences in that they differ from traditional sampling designs which address specific questions or hypotheses and emphasize adherence to statistical assumptions (e.g., randomness, independence, stationarity, and normality). In addition, conventional statistical methods generate inferences by considering the properties of small samples taken from a population. Big data have the potential to sample the majority of the population, reducing the need for tests of statistical significance.

The lack of a specific context and the massive sample size of big data create unique opportunities for scientific discovery, but pose significant challenges wherein high data volume and dimensionality can create noise accumulation, spurious correlations, and incidental endogeneity

(Fan et al. 2014, Gandomi and Haider 2015). The challenge of analyzing big data has largely been addressed through the development of big data analytics (Kambatla et al. 2014, Gandomi and Haider 2015) and cloud computing (Assunção et al. 2015, Hashem et al. 2015). These techniques and resources have allowed researchers to automatically and efficiently mine data using machine learning algorithms, which detect patterns, build predictive models, and optimize outcomes (Hastie et al. 2009, Han et al. 2011, Witten et al. 2016). The development of these methods and resources has promoted the rapid advancement of a new epistemological perspective in science in which knowledge is extracted directly from data (Kitchin 2014).

The emergence of big data in the natural sciences has opened novel lines of research, presented new challenges, and changed how we observe and study the natural world. Within ornithology, big data have begun to transform our knowledge about birds, from their annual distributions to how they relate to different environmental factors. To address this transformation, we provide an overview of the historical precursors to big data in ornithology and the past, present, and future of big data in ornithology. Our aim is to explore the rewards and risks associated with the application of big data and how this can best be used to advance the science of ornithology.

## Origins of Big Data in Ornithology

Ornithological big data did not appear overnight, but rather developed as both a data source and a means of inference over the past century. Big data precursors started with narrow purposes and goals, but became sources of broader inference through the expansion and repurposing of datasets to address emerging environmental challenges. Other big data sources were conceived as "big" from their outset, such as the citizen science project eBird (Sullivan et al. 2014). A third source of big data is data streams that exist outside ornithology—data that are collected and archived for other purposes but now provide unique research opportunities for ornithologists. This categorization based on purpose (ornithology-focused, -related, or -independent) is strongly correlated with a categorization based on data structure (structured, semistructured, and unstructured), as purpose-driven data collection generally results in systematically structured data. As the structure-based categorization has been broadly used in the big data literature (Gandomi and Haider 2015), we use it here to demarcate the current sources of big data in ornithology.

**Structured data.** Structured data are typically characterized by low volume, velocity, and variety. In ornithology, structured data have limits to what kinds of data are collected and how the data should be collected. Any data collection process that has or could have a "protocol," for example, is likely to result in structured data. As such, structured data are often collected within the context of a predefined question and tend to be highly organized based on space, time, taxa, and measurements, and changes to these limits and parameters are often carefully controlled by researchers. Because of these controls, most field research in ornithology produces structured data, with carefully conceived collection schemes that result in datasets of manageable size and complexity.

In ornithology, there is a variety of examples of structured datasets with large volumes and low variety and velocity, including the long-running national and regional bird population monitoring programs such as the North American Breeding Bird Survey (BBS), the British and French Breeding Bird Survey, and Audubon's Christmas Bird Count (CBC). These datasets can be identified as precursors to the more recent advent of semistructured and unstructured data in ornithology through their ability to provide a large volume of data annually over multiple decades. Ornithologists are increasingly using these resources within a broader and more innovative conceptual and analytical framework.

A prime example is the BBS, which has used a structured data collection scheme to acquire data on breeding bird populations across much of North America since 1966. Since its inception, the number of survey routes (39-km roadside surveys) in the BBS program has grown from $\sim$500 to $>$4,000, and the BBS now estimates abundance trends for more than 420 species (Link et al. 2017, Sauer et al. 2017). Currently, the BBS dataset contains more than 6.2 million independent observations of $>$730 bird species from across North America. Although the program tests alternative protocols for improving detectability and count estimation (Nichols et al. 2000), the sampling protocol for the primary point count data has remained relatively unchanged since inception. Raw data are compiled and hosted by the United States Geological Survey and are publicly available. Analytical methods for estimating population trends have evolved concurrently with advancements in technology and computational power, advances in ecological statistics, and with the increase in the length of the time series data (Sauer et al. 2017).

The BBS was initially created to monitor populations of songbirds and other nongame species (Robbins et al. 1986, 1989). The use of BBS data has now expanded far beyond the examination of population trends to include investigation of broad-scale conservation issues such as habitat loss and degradation and climate change (Flather and Sauer 1996, Lepczyk et al. 2008, Hudson et al. 2017) as well as the testing of biogeographical and ecological questions (Rowhani et al. 2008). Over time, the BBS has provided ecological inference far beyond its original goals, using analytical approaches for which the dataset was not designed. For example, the BBS has been used to explore questions related to species–energy theory (Dobson et al. 2015, Fristoe 2015), neutral theory (Kalyuzhny et al. 2014), the distinction between core and transient species in ecological communities (Coyle et al. 2013), and the spatial scaling of biotic interactions (Belmaker et al. 2015). It has additionally been used to build spatiotemporal niche models (Bateman et al. 2016a), to help make isotopic signature assignments (Hobson et al. 2014), and to study continental-scale population responses to urbanization (Pidgeon et al. 2014) and climate change (Stephens et al. 2016). The ever-growing volume of BBS data has provided opportunities for scientists to address hypotheses beyond the original scope of the monitoring program. In the case of the BBS, the advantages of structured data appeals to many ornithologists by offering strict standards and consistent methodologies, by utilizing stratified site locations to support representative sampling (Veech et al. 2017), and by providing the opportunity to track observer experience (Sauer et al. 1994). However, many of these same advantages that enhance data value and veracity can limit the form and breadth of scientific inquiry.

**Semistructured data.** Semistructured data are characterized by high volume, velocity, and variety, and lack the strict standards or methodologies of structured data. However, unlike unstructured data, semistructured data are organized in a fashion that more readily promotes analysis. The few examples of semistructured data are often considered to be a form of structured data. In ornithology, semistructured databases have arisen from efforts to increase the flexibility of traditional structured datasets in terms of sampling protocols and objectives. The data generated from these efforts tend to have lower veracity and do not conform to the more traditional sampling designs associated with structured data. The rise of semistructured data has coincided with the emergence of citizen science research agendas in ornithology (Dickinson et al. 2010, Cooper et al. 2014) that have led to a greater diversity of programs that are less prescriptive in their protocols, methodologies, and sampling designs. Citizen science programs also increase public awareness of scientific research, and some can contribute to social well-being (Bonney et al. 2016).

The primary example of semistructured big data in ornithology is the eBird citizen science database (Sullivan et al. 2014). eBird is a global bird monitoring project that allows volunteers to enter their observations of bird occurrence and abundance from any location at any time. Using several basic sampling protocols, observations are organized into a checklist format, which can then be entered into a central online data depository. Like other semistructured datasets, there are fewer requirements regarding sampling design, so eBird checklists include information on observer effort as defined by features of each sampling protocol. Importantly, eBird is semistructured in that while observers are required to use a checklist format within their chosen sampling protocol, they are free to collect data at any spatial and temporal resolution. Thus, observations compiled by eBird can range in quality and extent from very general to very detailed. By welcoming all types of data on bird occurrence and abundance, with only loose boundaries used to define quality and collection method, eBird has created a semistructured dataset with unprecedented volume and velocity. By mid-2017, since its inception in 2002, eBird had compiled over 30 million checklists containing more than 423 million observations. eBird data have been used to reveal patterns and determinants of broad-scale migration strategies (La Sorte et al. 2016a), and to advance our understanding of how migratory birds are associated across the annual cycle with protected areas and different land-cover categories (La Sorte et al. 2015a, Zuckerberg et al. 2016), nighttime light pollution (La Sorte et al. 2017b), and projected changes in climate and land use (La Sorte et al. 2017a).

**Unstructured data.** Unstructured data are characterized by extremely high volume, velocity, and variety. Unstructured data lack any intentional structure or organization and are often characterized by passive sensors that continuously collect text, images, audio, or video from the environment. In most cases, these efforts have no specific objective beyond data acquisition. In other cases, a nonornithological objective is present, but the data can be repurposed for ornithological research. More recently, ornithological questions have been used to define an underlying purpose for unstructured data acquisition.

A fundamental example of unstructured data that has been repurposed for ornithological research is data from weather surveillance radar (WSR). Since the 1940s, it has been known that radar can detect birds in flight (Lack and Varley 1945), and subsequently that it can be used to study multiple aspects of bird migration within the atmosphere (Eastwood 1967, Bruderer 1997a, 1997b). The advent of large networks of WSR stations in North America and Europe has allowed researchers to document migration patterns and associations within the atmosphere across broad geographic extents (Gauthreaux and Belser 1998). Current WSR systems are designed to monitor and track meteorological events, primarily precipitation. Extracting biological signals from WSR is challenging due to the volume of data generated by WSR and the overall complexity of the radar information where precipitation is detected in combination with other atmospheric contaminants such as insects, birds, bats, and dust. Nevertheless, several approaches have been developed using machine learning and other big data analytics to efficiently extract altitudinal profiles of bird density, speed, and direction from WSR images (Dokter et al. 2011, Farnsworth et al. 2016). This information has been used to assess how environmental factors such as wind speed and direction dictate migration timing (La Sorte et al. 2015b, Horton et al. 2016), and how nighttime light pollution affects local (Van Doren et al. 2017) and regional migratory behavior (McLaren et al. 2018). When WSR data are combined with modeled estimates of species' distributions using eBird observations, more refined ecological assessments can be made based on estimates of community composition at WSR stations across time (La Sorte et al. 2015b, 2015c).

A source of unstructured data that has been specifically designed to acquire ornithological information is noninvasive acoustic monitoring (Blumstein et al. 2011). Birds are a highly vocal taxa, and the collection of acoustic behaviors using microphone arrays has been used to monitor bird populations and their behavior during stationary periods (Acevedo and Villanueva-Rivera 2006, Dawson and Efford 2009, Vallejo and Taylor 2009) and during migration (Farnsworth 2005). The resulting acoustic information has been used to estimate species richness (Depraetere et al. 2012, Wimmer et al. 2013) and monitor changes in species density (Dawson and Efford 2009) and community composition over time (Lellouch et al. 2014). During migration, many bird species, especially nocturnal migrants, emit short vocalizations during flight. These flight calls can be used to identify species and provide information on species composition and behavior during migration (Watson et al. 2016). Likewise, autonomous recording units have been used to detect the presence and spatial distribution of a variety of seabird species (Buxton and Jones 2012, Cragg et al. 2015, Harvey et al. 2016). In contrast to manual techniques for species identification, the development of automated machine learning methodologies (Bardeli et al. 2010, Digby et al. 2013, de Oliveira et al. 2015, Stowell et al. 2016, Zhao et al. 2017) has the potential to rapidly advance the use of unstructured acoustic monitoring in ornithological research (Gorrepati et al. 2012).

## Past and Present Contributions to Ornithology

The flexible nature of semistructured and unstructured data has allowed ornithologists to address novel questions and test long-standing hypotheses using unique biological perspectives and observational scales. Unlike annual monitoring programs that are confined to particular time periods, such as the breeding season, semistructured or unstructured programs allow continuous and flexible data coverage and the ability to collect a greater diversity of data on bird populations. Here, we describe how these data have been instrumental in documenting novel patterns and associations and their role in advancing our knowledge of avian distributions at unique spatial and temporal scales. Key benefits include the ability to address questions at the level of entire populations across the full annual cycle and the full geographic extent of a species' annual distribution. These efforts are also readily scalable, allowing researchers to address questions across multiple taxa or for entire species assemblages.

**Diversity patterns.** Understanding patterns of biodiversity at geographic scales is one of several research areas that could only be addressed well with the advent of large datasets. Data from the BBS and CBC have shed light on continental-scale patterns of diversity during the breeding and nonbreeding seasons (Hurlbert and Haskell 2003) and have revealed important differences from what would be expected simply by overlaying range maps (Hurlbert and White 2005). These datasets have facilitated the testing of species–energy theory as a driver of richness patterns (Hurlbert 2004, Dobson et al. 2015) and the relative roles of local and regional processes (White and Hurlbert 2010, Coyle et al. 2013) and interannual variation in climate (Rowhani et al. 2008) as drivers of those patterns. To date, the majority of diversity research has been accomplished using structured datasets, but some of the newer semistructured and unstructured data sources will undoubtedly allow more refined examinations of diversity patterns at larger extents and finer grains across the full annual cycle.

**Species distribution modeling.** The occurrence and abundance information contained in eBird and other semistructured programs does not lend itself to traditional forms of parametric or nonparametric analysis to estimate where birds occur across the annual cycle. The dynamic nature of bird distributions requires approaches that can accommodate the presence of spatiotemporal variation within and across scales. For birds, these dynamics are dependent on the phase of the annual cycle, with stationary periods (breeding and nonbreeding) providing greater structure, and more dynamic periods (migration) creating greater heterogeneity. Therefore, a data-driven approach is needed that can estimate distributions without having to model the underlying dynamic processes, which is often required when using traditional analytical approaches. This can be achieved, for example, by using an ensemble or mixture model approach, which implements a large number of static species distribution models each applied to a spatiotemporally restricted extent whose form adapts to spatiotemporal variation in data density (Fink et al. 2010, 2014). Predictions can then be generated by averaging across local models with shared extents, allowing local patterns to scale up to estimate patterns at regional scales (Fink et al. 2010, 2014). An alternative to modeling species distributions using semistructured data alone is to combine structured and semistructured data, with the goal of balancing the tradeoffs between data quality and quantity to improve model breadth and performance (Fithian et al. 2015, Giraud et al. 2016, Fournier et al. 2017, Pacifici et al. 2017). These data fusion approaches are particularly valuable when estimating the distributions of very rare species or species that are difficult to detect, for which the quantity and quality of structured data are often severely lacking (Fournier et al. 2017).

**Broad-scale migration strategies.** The advent of big data has created unique opportunities to study bird migration from a macroecological perspective. Here, broad-scale patterns and associations can be documented, and questions and hypotheses can be addressed or tested in a taxonomically and geographically comprehensive fashion. For example, using observations from the eBird database, broad-scale migration strategies can be documented for multiple taxa at the population level across the full annual cycle. Such work has provided evidence that many migratory bird populations employ looped migration strategies (La Sorte et al. 2016a). By testing alternative migration scenarios, it has also been shown that seasonal variation in atmospheric conditions (La Sorte et al. 2014b) and ecological productivity (La Sorte et al. 2014a) have played primary roles in promoting the development of these migration strategies. Researchers can now formulate questions to address sources of variation within these broad-scale commonalities, such as migration distances or the presence of ecological barriers to migration (La Sorte and Fink 2017).

**Migration and breeding phenology.** Big data are contributing to ornithology where measured phenomena happen over short periods of time with poor predictability. The measurement of such phenomena has historically required massive amounts of effort from scientists for little data, whereas with big data researchers have the power to monitor these phenomena nearly continuously with little direct effort. For this reason, big data are greatly advancing our understanding of avian phenology, including the timing of migration and breeding phenophases. Some big data resources, such as the citizen science project Journey North, focus exclusively on aspects of phenology such as migration arrival (Arab et al. 2016). Likewise, eBird is increasingly being used to understand migration phenology, particularly how the timing of migration has been shifting in response to recent climate change, and the extent to which these shifts vary geographically and with species traits (Hurlbert and Liang 2012, Mayor et al. 2017). Breeding phenophases (e.g., nest building, egg laying, egg hatching, and fledging) have proven more difficult to measure, with the exception of the dedicated citizen science project NestWatch (Cooper 2014). There are many opportunities for advances in extracting phenological signals from passive sensors in the future, such as estimating the timing of breeding activities based on the frequency of passively recorded songbird vocalizations (Strebel et al. 2014).

**Demography.** Avian demographic research has played an important role in advancing the field of avian population ecology (Sæther and Bakke 2000, Sillett and Holmes 2002) and has played a critical role in supporting bird conservation (Green 1999). These efforts have taken on greater relevance for migratory bird species as researchers explore the demographic drivers underlying current population declines (Morrison et al. 2016, Border et al. 2017). Several large datasets with highly structured protocols provide a unique platform for estimating demographic parameters and population trends across broad spatial and temporal extents. For example, long-term ringing or banding programs, such as the North American Monitoring Avian Productivity and Survivorship (MAPS) program (Saracco et al. 2010, 2012) and the British Trust for Ornithology's (BTO) Constant Effort Sites (CES) scheme, have allowed researchers to develop more comprehensive and detailed models of population change and viability (Cave et al. 2010, Ahrestani et al. 2017).

**Geographic range shifts.** For decades, geographic ranges of birds were represented by static range maps residing in field guides and largely based on expert opinion. Big data have opened the door to displaying geographic ranges in a more dynamic fashion that fully captures the temporal complexity of bird distributions. These efforts are promoting new avenues of research. For example, under global warming, geographic range boundaries of birds and other taxa are responding by shifting to higher latitudes (Chen et al. 2011). Researchers have relied on structured datasets collected over many decades to document these shifts for breeding and wintering bird populations in North America and Europe (Hitch and Leberg 2007, La Sorte and Thompson 2007, Mason et al. 2015). In recent years, these datasets have shown that the patterns and drivers of range shifts are considerably more variable than initially thought. For example, geographic range shifts of variable directions and intensities have been documented for birds in North America (Bateman et al. 2016b), Great Britain (Gillings et al. 2015), and Australia (VanDerWal et al. 2013), and there is evidence that the geographic responses of some species contain lag effects that can encompass several decades (La Sorte and Jetz 2012). In addition, there is evidence that range shifts cannot be accurately predicted by species' traits (Angert et al. 2011) and that changes in climatic factors such as precipitation (McCain and Colwell 2011, Tingley et al. 2012) or the frequency and intensity of climatic extremes (La Sorte et al. 2016b) can play a significant role determining geographic responses. As the temporal extent of semistructured and unstructured data continues to grow, researchers will be poised to document how birds respond geographically to global warming with greater temporal and spatial detail, further elucidating the primary trends and their drivers.

## Challenges of Big Data Ornithology

**Statistical issues.** The massive volume and high dimensionality of big data have promoted the development of new statistical and computational methods, but significant challenges for analysis and interpretation remain. These challenges are often characterized as noise accumulation, spurious correlations, and incidental endogeneity (Fan et al. 2014, Gandomi and Haider 2015).

A common feature of big data is high dimensionality, in which many variables are measured across a large number of samples. Applying traditional parametric approaches to these datasets can be problematic when individual variables within the dataset can be accurately predicted by linear combinations of other variables (multicollinearity). Under these circumstances, coefficient estimates can respond erratically to small changes in the data or the statistical model. Methods do exist that allow for robust statistical inference when the numbers of parameters in a dataset are exceptionally large (Bühlmann and van de Geer 2011). Big data resources in ornithology typically contain few variables, but this can quickly change if ornithological data is combined with highly dimensional environmental datasets.

In addition to issues related to multicollinearity, when estimating or testing many parameters simultaneously, the accumulation of noise or errors has the potential to mask variables that have true effects, an issue that can intensify as data dimensionality increases (Fan et al. 2014). Sparse models and variable selection can overcome issues related to noise accumulation. However, variable selection using highly dimensional data is also affected by spurious correlations, incidental endogeneity, heterogeneity, and measurement errors, compounding the challenges of generating robust inferences (Fan et al. 2014).

Spurious correlation refers to uncorrelated variables that are falsely classified as being correlated due to the extreme size of the dataset. As shown by Fan et al. (2014), the statistical significance of the correlation between 2 independent variables tends to increase as data volume increases, which can result in spurious correlations and erroneous conclusions. When analyzing large datasets, incidental endogeneity is often present when the residual term is dependent on some of the predictors. This exogenous assumption (i.e. the independence of the residuals and predictors) is central to most statistical methods, and, unlike spurious correlations, this assumption is violated when a genuine statistical relationship exists.

In sum, understanding the statistical challenges associated with extremely large sample sizes and high dimensionality is central to working with big data. To effectively handle these challenges, methods designed to address data complexity, noise, and data dependencies are needed (Fan et al. 2014). Many of these methods are available or are being developed, but researchers still need to apply these methods with a clear understanding of existing limitations and how they can affect resulting inferences. Inferential quality can be further enhanced by using theory and process to guide data selection and analysis, and by avoiding open research questions that rely on exploring data for correlations (Coveney et al. 2016).

**Data veracity.** With increased flexibility and less structure comes substantial variability in data quality, which is identified as data veracity. In response to the concern of reduced precision and increased uncertainty, some big data ornithology programs have developed sophisticated processes of data validation. For example, Project FeederWatch and eBird use a combination of automated filters (Bonter and Cooper 2012) and expert review (Kelling et al. 2013) to assess the veracity of observations. The interaction between these 2 components allows continued improvement in data quality through the refinement of data filters that flag questionable records for expert review. Ancillary data that estimate variation in sampling effort (e.g., survey duration, transect length, number of observers) can also be collected. Researchers can use this information to create subsets of the data to achieve the level of uncertainty or quality required to address a specific research question or objective, often without a significant loss of data volume. Ancillary information can also be included as covariates in statistical models, allowing researchers to standardize effort to a common baseline when implementing analyses (Fink et al. 2010, 2014). Despite these approaches for enhancing data veracity, there are still lingering issues around whether these efforts are sufficient. Issues of trust, whether grounded in reality or not, are likely to remain a challenge for scientists seeking to use these resources. In particular, the role of data veracity (especially false positives) will be of express concern to users of unstructured data, where data are generated with very limited filtering.

## Future Contributions of Big Data to Ornithology

**Crowdsourced ornithological data.** Data on birds can come from nontraditional sources to which users contribute either accidental or nontraditional ornithological data online. A semistructured example would be the xeno-canto database (www.xeno-canto.org), to which users upload and identify avian vocalizations. With >350,000 recordings, xeno-canto provides opportunities to study phenology and behavior beyond those available using typical, structured vocalization datasets. Perhaps more intriguing, however, is the potential contribution from unstructured ornithological crowdsourcing via online photography repositories (Leighton et al. 2016) or social media (e.g., Twitter or Facebook). Such web platforms are already used for crowdsourcing environmental and public health information (Kamel Boulos et al. 2011, Alvaro et al.

2015), and hold untapped potential for ornithological research, particularly when human–bird interactions are of interest.

**Tracking migratory birds.** Individual tracking technology has become increasingly refined over the past several decades, providing opportunities for rapid advancements in the breadth and quantity of data that can be acquired on individual migrating birds (Lopez-Lopez 2016). Geolocators (Stutchbury et al. 2009) and, more recently, satellite transmitters are providing increasingly more detailed information on where migratory birds occur across space and time, including information on flight speed and altitude in some cases (Bridge et al. 2011). In addition to providing movement information, other sensors are becoming available that can provide ancillary information on behavior or characteristics of the surrounding environment (Wilmers et al. 2015). Other advancements in individual tracking technology include the Motus Wildlife Tracking System (https://motus.org), which uses an automated array of detectors that can be positioned across a broad geographic region (Taylor et al. 2017). Data repositories such as Movebank (http://www.movebank.org/) have also facilitated the storage and retrieval of tracking information, which is likely to take on greater relevance as the velocity and volume of tracking data expands.

**Autonomous recording units.** Traditional field ornithology has relied on survey methods such as point counts to monitor bird populations. The ongoing development and implementation of autonomous recording units (ARUs) are replacing these traditional methods by sampling the acoustics of bird communities. ARUs can provide repeated data collection over time, reduce the potential for observer bias, allow sampling of many locations (including remote or hard-to-access sites), and provide a permanent record of the survey (Shonfield and Bayne 2017). While ARUs can be very useful for surveys in remote locations, especially of rare, cryptic, or secretive species (Drake et al. 2016), and for general biodiversity sampling (Shonfield and Bayne 2017), the consistent methodology and ability to collect data over long periods of time allow acoustic data to be integrated with other data sources. For instance, information from ARUs can be used to model bird occupancy over time across broad geographic regions (Furnas and Callas 2015). Given the decreasing cost of ARU technology, ARUs may offer new and expanding research opportunities.

**Real-time ecological assessments.** Several big data resources have the potential to provide real-time information on the evolutionary and ecological implications of rapid environmental change (La Salle et al. 2016). Observations compiled by citizen scientists can play a key role in this process (Bonney et al. 2009, Dickinson et al. 2010). For example, estimates of bird occurrence or abundance (eBird) or of migration intensity (WSR) can be used to address how bird populations are affected by extreme weather events (e.g., heat waves, droughts, tornadoes, and hurricanes; Albright et al. 2010a, 2010b) or human-caused natural disasters (e.g., oil spills or wildfires). One example that represents the real-time potential of these efforts is a study that examined the climatic drivers of Pine Siskin (*Spinus pinus*) irruptive migration in North America (Strong et al. 2015). This study used 2 million Pine Siskin observations from the Project FeederWatch citizen science program (https://feederwatch.org/), which monitors the occurrence and abundance of wintering birds at supplemental feeders in North America, to assess how irruptions were correlated with climatic variability. A second example is a study that used eBird occurrence information to document how migration phenology and breeding season occurrence for 353 North American bird species were affected by an extreme warming event that occurred during spring migration (La Sorte et al. 2016b). By advancing the quality and efficiency of the methods used in these studies, there is the potential to generate detailed and rigorous real-time assessments of the implications of extreme events or natural disasters for bird populations, which can be used to inform conservation efforts on the ground as well as long-term mitigation strategies.

**Microscale behavioral ecology.** Big data do not necessarily need to be collected over big scales. Radio-Frequency Identification (RFID) technology dates back to the 1970s, but in recent years ornithologists have used RFID to examine questions related to feeding rates, incubation behavior, changes in body condition, movement, dispersal, and social networks (see review by Bonter and Bridge 2011). RFID-based studies are local and focus on a limited number of species. The approach relies on fastening a Passive Integrated Transponder (PIT) to the leg band of a bird, and then equipping feeding stations with an RFID data logger. Each time a PIT-tagged bird lands on a feeder, its unique identification number and the date and time of the visit are recorded. Given that these studies include hundreds of individual birds, the resulting databases can be extremely large and complex. For example, a study conducted during a single winter season in central New York, USA, generated a database consisting of more than 450,000 feeder visits, including $>$200 visits per day by a single individual (Bonter et al. 2013). Many of these RFID networks are increasing in scale and scope, wherein hundreds of RFID readers are collecting data continuously on individual behavior. The use of RFID has become so popular that an R package is available for managing and graphing data collected from RFID-equipped feeders (LaZerte 2017).

## Conclusions

Big data resources and computational and analytical techniques and tools provide singular opportunities for researchers to describe unique properties of complex natural systems (Bar-Yam 2016), which in some cases may complement insights generated using traditional observational or experimental approaches (Kelling et al. 2009). Big data have contributed and continue to add empirical breadth and detail to many scientific disciplines, but these resources must be used with a clear understanding of their limitations. Thus, in addition to developing ornithological applications, effort is needed to better understand how these limitations can be mitigated through refined analytical methods or study designs, and how scientific questions and hypotheses can be best formulated to maximize inferential quality and rigor.

## ACKNOWLEDGMENTS

We thank the North American Ornithological Conference for hosting our 2016 symposium on big data ornithology. In addition, we are grateful to J. Nichols and 2 anonymous reviewers for feedback on the manuscript.

**Ethics statement:** This is a review article and does not include new data.

**Author contributions:** C.A.L. proposed the symposium and conceived the idea for the review article. All authors contributed equally to the conception, literature review, and writing of the paper. Author order reflects the 3 organizers of the symposium, followed by symposium participants in alphabetical order.

## LITERATURE CITED

Acevedo, M. A., and L. J. Villanueva-Rivera (2006). Using automated digital recording systems as effective tools for the monitoring of birds and amphibians. Wildlife Society Bulletin 34:211–214.

Ahrestani, F. S., J. F. Saracco, J. R. Sauer, K. L. Pardieck, and J. A. Royle (2017). An integrated population model for bird monitoring in North America. Ecological Applications 27: 916–924.

Albright, T. P., A. M. Pidgeon, C. D. Rittenhouse, M. K. Clayton, C. H. Flather, P. D. Culbert, B. D. Wardlow, and V. C. Radeloff (2010a). Effects of drought on avian community structure. Global Change Biology 16:2158–2170.

Albright, T. P., A. M. Pidgeon, C. D. Rittenhouse, M. K. Clayton, B. D. Wardlow, C. H. Flather, P. D. Culbert, and V. C. Radeloff (2010b). Combined effects of heat waves and droughts on avian communities across the conterminous United States. Ecosphere 1:art.12.

Alvaro, N., M. Conway, S. Doan, C. Lofi, J. Overington, and N. Collier (2015). Crowdsourcing Twitter annotations to identify first-hand experiences of prescription drug use. Journal of Biomedical Informatics 58:280–287.

Angert, A. L., L. G. Crozier, L. J. Rissler, S. E. Gilman, J. J. Tewksbury, and A. J. Chunco (2011). Do species' traits predict recent shifts at expanding range edges? Ecology Letters 14: 677–689.

Arab, A., J. R. Courter, and J. Zelt (2016). A spatio-temporal comparison of avian migration phenology using citizen science data. Spatial Statistics 18:234–245.

Assunção, M. D., R. N. Calheiros, S. Bianchi, M. A. S. Netto, and R. Buyya (2015). Big data computing and clouds: Trends and future directions. Journal of Parallel and Distributed Computing 79:3–15.

Bardeli, R., D. Wolff, F. Kurth, M. Koch, K. H. Tauchert, and K. H. Frommolt (2010). Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring. Pattern Recognition Letters 31:1524–1534.

Bar-Yam, Y. (2016). From big data to important information. Complexity 21:73–98.

Bateman, B. L., A. M. Pidgeon, V. C. Radeloff, C. H. Flather, J. VanDerWal, H. R. Akçakaya, W. E. Thogmartin, T. P. Albright, S. J. Vavrus, and P. J. Heglund (2016a). Potential breeding distributions of U.S. birds predicted with both short-term variability and long-term average climate data. Ecological Applications 26:2720–2731.

Bateman, B. L., A. M. Pidgeon, V. C. Radeloff, J. VanDerWal, W. E. Thogmartin, S. J. Vavrus, and P. J. Heglund (2016b). The pace of past climate change vs. potential bird distributions and land use in the United States. Global Change Biology 22: 1130–1144.

Belmaker, J., P. Zarnetske, M.-N. Tuanmu, S. Zonneveld, S. Record, A. Strecker, and L. Beaudrot (2015). Empirical evidence for the scale dependence of biotic interactions. Global Ecology and Biogeography 24:750–761.

Benson, B. J., B. J. Bond, M. P. Hamilton, R. K. Monson, and R. Han (2010). Perspectives on next-generation technology for environmental sensor networks. Frontiers in Ecology and the Environment 8:193–200.

Blumstein, D. T., D. J. Mennill, P. Clemins, L. Girod, K. Yao, G. Patricelli, J. L. Deppe, A. H. Krakauer, C. Clark, K. A. Cortopassi, S. F. Hanser, et al. (2011). Acoustic monitoring in terrestrial environments using microphone arrays: Applications, technological considerations and prospectus. Journal of Applied Ecology 48:758–767.

Bonney, R., C. B. Cooper, J. Dickinson, S. Kelling, T. Phillips, K. V. Rosenberg, and J. Shirk (2009). Citizen science: A developing tool for expanding science knowledge and scientific literacy. Bioscience 59:977–984.

Bonney, R., T. B. Phillips, H. L. Ballard, and J. W. Enck (2016). Can citizen science enhance public understanding of science? Public Understanding of Science 25:2–16.

Bonter, D. N., and E. S. Bridge (2011). Applications of radio frequency identification (RFID) in ornithological research: A review. Journal of Field Ornithology 82:1–10.

Bonter, D. N., and C. B. Cooper (2012). Data validation in citizen science: A case study from Project FeederWatch. Frontiers in Ecology and the Environment 10:305–307.

Bonter, D. N., B. Zuckerberg, C. W. Sedgwick, and W. M. Hochachka (2013). Daily foraging patterns in free-living birds: Exploring the predation–starvation trade-off. Proceedings of the Royal Society B 280:20123087.

Border, J. A., I. G. Henderson, D. Ash, and I. R. Hartley (2017). Characterising demographic contributions to observed population change in a declining migrant bird. Journal of Avian Biology 48:1139–1149.

Bowker, G. C. (2000). Biodiversity datadiversity. Social Studies of Science 30:643–683.

Boyd, D., and K. Crawford (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. Information, Communication & Society 15: 662–679.

Bridge, E. S., K. Thorup, M. S. Bowlin, P. B. Chilson, R. H. Diehl, R. W. Fléron, P. Hartl, R. Kays, J. F. Kelly, W. D. Robinson, and M. Wikelski (2011). Technology on the move: Recent and forthcoming innovations for tracking migratory birds. Bioscience 61:689–698.

Bruderer, B. (1997a). The study of bird migration by radar, Part 1: The technical basis. Naturwissenschaften 84:1–8.

Bruderer, B. (1997b). The study of bird migration by radar, Part 2: Major achievements. Naturwissenschaften 84:45–54.

Bühlmann, P., and S. van de Geer (2011). Statistics for High-Dimensional Data: Methods, Theory and Applications. Springer-Verlag, Heidelberg, Germany.

Buxton, R. T., and I. L. Jones (2012). Measuring nocturnal seabird activity and status using acoustic recording devices: Applications for island restoration. Journal of Field Ornithology 83: 47–60.

Cave, V. M., R. King, and S. N. Freeman (2010). An integrated population model from constant effort bird-ringing data. Journal of Agricultural, Biological, and Environmental Statistics 15:119–137.

Chen, I.-C., J. K. Hill, R. Ohlemüller, D. B. Roy, and C. D. Thomas (2011). Rapid range shifts of species associated with high levels of climate warming. Science 333:1024–1026.

Cooper, C. B. (2014). Is there a weekend bias in clutch-initiation dates from citizen science? Implications for studies of avian breeding phenology. International Journal of Biometeorology 58:1415–1419.

Cooper, C. B., J. Shirk, and B. Zuckerberg (2014). The invisible prevalence of citizen science in global research: Migratory birds and climate change. PLOS One 9:e106508.

Coveney, P. V., E. R. Dougherty, and R. R. Highfield (2016). Big data need big theory too. Philosophical Transactions of the Royal Society A 374:20160153.

Coyle, J. R., A. H. Hurlbert, and E. P. White (2013). Opposing mechanisms drive richness patterns of core and transient bird species. The American Naturalist 181:E83–E90.

Cragg, J. L., A. E. Burger, and J. F. Piatt (2015). Testing the effectiveness of automated acoustic sensors for monitoring vocal activity of Marbled Murrelets *Brachyramphus marmoratus*. Marine Ornithology 43:151–160.

Dawson, D. K., and M. G. Efford (2009). Bird population density estimated from acoustic signals. Journal of Applied Ecology 46:1201–1209.

de Oliveira, A. G., T. M. Ventura, T. D. Ganchev, J. M. de Figueiredo, O. Jahn, M. I. Marques, and K.-L. Schuchmann (2015). Bird acoustic activity detection based on morphological filtering of the spectrogram. Applied Acoustics 98:34–42.

Depraetere, M., S. Pavoine, F. Jiguet, A. Gasc, S. Duvail, and J. Sueur (2012). Monitoring animal diversity using acoustic indices: Implementation in a temperate woodland. Ecological Indicators 13:46–54.

Devictor, V., and B. Bensaude-Vincent (2016). From ecological records to big data: The invention of global biodiversity. History and Philosophy of the Life Sciences 38:art.13.

Dickinson, J. L., B. Zuckerberg, and D. N. Bonter (2010). Citizen science as an ecological research tool: Challenges and benefits. Annual Review of Ecology, Evolution, and Systematics 41:149–172.

Digby, A., M. Towsey, B. D. Bell, and P. D. Teal (2013). A practical comparison of manual and autonomous methods for acoustic monitoring. Methods in Ecology and Evolution 4: 675–683.

Dobson, L. L., F. A. La Sorte, L. L. Manne, and B. A. Hawkins (2015). The diversity and abundance of North American bird assemblages fail to track changing productivity. Ecology 96: 1105–1114.

Dokter, A. M., F. Liechti, H. Stark, L. Delobbe, P. Tabary, and I. Holleman (2011). Bird migration flight altitudes studied by a network of operational weather radars. Journal of the Royal Society Interface 8:30–43.

Drake, K. L., M. Frey, D. Hogan, and R. Hedley (2016). Using digital recordings and sonogram analysis to obtain counts of Yellow Rails. Wildlife Society Bulletin 40:346–354.

Eastwood, E. (1967). Radar Ornithology. Methuen, London, UK.

Fan, J., F. Han, and H. Liu (2014). Challenges of big data analysis. National Science Review 1:293–314.

Farnsworth, A. (2005). Flight calls and their value for future ornithological studies and conservation research. The Auk 122:733–746.

Farnsworth, A., B. M. Van Doren, W. M. Hochachka, D. Sheldon, K. Winner, J. Irvine, J. Geevarghese, and S. Kelling (2016). A characterization of autumn nocturnal migration detected by weather surveillance radars in the northeastern USA. Ecological Applications 26:752–770.

Fink, D., T. Damoulas, N. E. Bruns, F. A. La Sorte, W. M. Hochachka, C. P. Gomes, and S. Kelling (2014). Crowdsourcing meets ecology: Hemispherewide spatiotemporal species distribution models. AI Magazine 35:19–30.

Fink, D., W. M. Hochachka, B. Zuckerberg, D. W. Winkler, B. Shaby, M. A. Munson, G. Hooker, M. Riedewald, D. Sheldon, and S. Kelling (2010). Spatiotemporal exploratory models for broad-scale survey data. Ecological Applications 20:2131–2147.

Fithian, W., J. Elith, T. Hastie, and D. A. Keith (2015). Bias correction in species distribution models: Pooling survey and collection data for multiple species. Methods in Ecology and Evolution 6:424–438.

Flather, C. H., and J. R. Sauer (1996). Using landscape ecology to test hypotheses about large-scale abundance patterns in migratory birds. Ecology 77:28–35.

Fournier, A. M. V., A. R. Sullivan, J. K. Bump, M. Perkins, M. C. Shieldcastle, and S. L. King (2017). Combining citizen science species distribution models and stable isotopes reveals migratory connectivity in the secretive Virginia Rail. Journal of Applied Ecology 54:618–627.

Fristoe, T. S. (2015). Energy use by migrants and residents in North American breeding bird communities. Global Ecology and Biogeography 24:406–415.

Furnas, B. J., and R. L. Callas (2015). Using automated recorders and occupancy models to monitor common forest birds across a large geographic region. The Journal of Wildlife Management 79:325–337.

Gandomi, A., and M. Haider (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management 35:137–144.

Gauthreaux, S. A., Jr., and C. G. Belser (1998). Displays of bird movements on the WSR-88D: Patterns and quantification. Weather and Forecasting 13:453–464.

Gillings, S., D. E. Balmer, and R. J. Fuller (2015). Directionality of recent bird distribution shifts and climate change in Great Britain. Global Change Biology 21:2155–2168.

Giraud, C., C. Calenge, C. Coron, and R. Julliard (2016). Capitalizing on opportunistic data for monitoring relative abundances of species. Biometrics 72:649–658.

Gorrepati, R. R., D.-H. Park, and D.-H. Kim (2012). A semantic sensing information representation for bird ecology. In Embedded and Multimedia Computing Technology and Service: EMC 2012 (J. J. Park, Y.-S. Jeong, S. O. Park, and H.-C. Chen, Editors). Springer, Dordrecht, Netherlands. pp. 429–433.

Green, R. E. (1999). Applications of large-scale studies of demographic rates to bird conservation. Bird Study 46: S279–S288.

Hampton, S. E., C. A. Strasser, J. J. Tewksbury, W. K. Gram, A. E. Budden, A. L. Batcheller, C. S. Duke, and J. H. Porter (2013). Big data and the future of ecology. Frontiers in Ecology and the Environment 11:156–162.

Han, J., M. Kamber, and J. Pei (2011). Data Mining: Concepts and Techniques, third edition. Morgan Kaufmann/Elsevier, Waltham, MA, USA.

Harvey, A. L., D. M. Mazurkiewicz, M. W. McKnown, K. W. Barnes, and M. W. Parker (2016). Changing breeding status of the Ashy Storm-Petrel *Oceanodroma homochroa* on Anacapa Island, California. Marine Ornithology 44:93–97.

Hashem, I. A. T., I. Yaqoob, Badrul Anuar, N. S. Mokhtar, A. Gani, and S. Ullah Khan (2015). The rise of "big data" on cloud computing: Review and open research issues. Information Systems 47:98–115.

Hastie, T., R. Tibshirani, and J. Friedman (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction, second edition. Springer-Verlag, New York, NY, USA.

Hitch, A. T., and P. L. Leberg (2007). Breeding distributions of North American bird species moving north as a result of climate change. Conservation Biology 21:534–539.

Hobson, K. A., S. L. Van Wilgenburg, J. Faaborg, J. D. Toms, C. Rengifo, A. L. Sosa, Y. Aubry, and R. Brito Aguilar (2014). Connecting breeding and wintering grounds of Neotropical migrant songbirds using stable hydrogen isotopes: A call for an isotopic atlas of migratory connectivity. Journal of Field Ornithology 85:237–257.

Horton, K. G., B. M. Van Doren, P. M. Stepanian, W. M. Hochachka, A. Farnsworth, and J. F. Kelly (2016). Nocturnally migrating songbirds drift when they can and compensate when they must. Scientific Reports 6:21249.

Hudson, M.-A. R., C. M. Francis, K. J. Campbell, C. M. Downes, A. C. Smith, and K. L. Pardieck (2017). The role of the North American Breeding Bird Survey in conservation. The Condor: Ornithological Applications 119:526–545.

Hurlbert, A. H. (2004). Species–energy relationships and habitat complexity in bird communities. Ecology Letters 7:714–720.

Hurlbert, A. H., and J. P. Haskell (2003). The effect of energy and seasonality on avian species richness and community composition. The American Naturalist 161:83–97.

Hurlbert, A. H., and Z. Liang (2012). Spatiotemporal variation in avian migration phenology: Citizen science reveals effects of climate change. PLOS One 7:e31662.

Hurlbert, A. H., and E. P. White (2005). Disparity between range map- and survey-based analyses of species richness: Patterns, processes and implications. Ecology Letters 8:319–327.

Jensen, J. R. (2006). Remote Sensing of the Environment: An Earth Resource Perspective, second edition. Pearson, London, UK.

Jin, X., B. W. Wah, X. Cheng, and Y. Wang (2015). Significance and challenges of big data research. Big Data Research 2:59–64.

Kalyuzhny, M., E. Seri, R. Chocron, C. H. Flather, R. Kadmon, and N. M. Shnerb (2014). Niche versus neutrality: A dynamical analysis. The American Naturalist 184:439–446.

Kambatla, K., G. Kollias, V. Kumar, and A. Grama (2014). Trends in big data analytics. Journal of Parallel and Distributed Computing 74:2561–2573.

Kamel Boulos, M. N., B. Resch, D. N. Crowley, J. G. Breslin, G. Sohn, R. Burtner, W. A. Pike, E. Jezierski, and K.-Y. Slayer Chuang (2011). Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management: Trends, OGC standards and application examples. International Journal of Health Geographics 10:art.67.

Keller, M., D. S. Schimel, W. W. Hargrove, and F. M. Hoffman (2008). A continental strategy for the National Ecological Observatory Network. Frontiers in Ecology and the Environment 6:282–284.

Kelling, S., J. Gerbracht, D. Fink, C. Lagoze, W.-K. Wong, J. Yu, T. Damoulas, and C. P. Gomes (2013). eBird: A human/computer learning network for biodiversity conservation and research. In Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence. AAAI Press, Toronto, ON, Canada. pp. 2229–2236.

Kelling, S., W. M. Hochachka, D. Fink, M. Riedewald, R. Caruana, G. Ballard, and G. Hooker (2009). Data-intensive science: A new paradigm for biodiversity studies. Bioscience 59:613–620.

Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. Big Data & Society 1. doi:10.1177/2053951714528481

Lack, D., and G. C. Varley (1945). Detection of birds by radar. Nature 156:446.

LaDeau, S. L., B. A. Han, E. J. Rosi-Marshall, and K. C. Weathers (2017). The next decade of big data in ecosystem science. Ecosystems 20:274–283.

Laney, D. (2001). 3-D data management: Controlling data volume, velocity and variety. META Group Research Note 6.

La Salle, J., K. J. Williams, and C. Moritz (2016). Biodiversity analysis in the digital era. Philosophical Transactions of the Royal Society B 371:20150337.

La Sorte, F. A., and D. Fink (2017). Migration distance, ecological barriers and en-route variation in the migratory behaviour of terrestrial bird populations. Global Ecology and Biogeography 26:216–227.

La Sorte, F. A., and W. Jetz (2012). Tracking of climatic niche boundaries under recent climate change. Journal of Animal Ecology 81:914–925.

La Sorte, F. A., and F. R. Thompson, III (2007). Poleward shifts in winter ranges of North American birds. Ecology 88:1803–1812.

La Sorte, F. A., D. Fink, P. J. Blancher, A. D. Rodewald, V. Ruiz-Gutierrez, K. V. Rosenberg, W. M. Hochachka, P. H. Verburg, and S. Kelling (2017a). Global change and the distributional

dynamics of migratory bird populations wintering in Central America. Global Change Biology 23:5284–5296.
La Sorte, F. A., D. Fink, J. J. Buler, A. Farnsworth, and S. A. Cabrera-Cruz (2017b). Seasonal associations with urban light pollution for nocturnally migrating bird populations. Global Change Biology 23:4609–4619.
La Sorte, F. A., D. Fink, W. M. Hochachka, J. L. Aycrigg, K. V. Rosenberg, A. D. Rodewald, N. E. Bruns, A. Farnsworth, B. L. Sullivan, C. Wood, and S. Kelling (2015a). Documenting stewardship responsibilities across the annual cycle for birds on U.S. public lands. Ecological Applications 25:39–51.
La Sorte, F. A., D. Fink, W. M. Hochachka, J. P. DeLong, and S. Kelling (2014a). Spring phenology of ecological productivity contributes to the use of looped migration strategies by birds. Proceedings of the Royal Society B 281:20140984.
La Sorte, F. A., D. Fink, W. M. Hochachka, A. Farnsworth, A. D. Rodewald, K. V. Rosenberg, B. L. Sullivan, D. W. Winkler, C. Wood, and S. Kelling (2014b). The role of atmospheric conditions in the seasonal dynamics of North American migration flyways. Journal of Biogeography 41:1685–1696.
La Sorte, F. A., D. Fink, W. M. Hochachka, and S. Kelling (2016a). Convergence of broad-scale migration strategies in terrestrial birds. Proceedings of the Royal Society B 283:20152588.
La Sorte, F. A., W. M. Hochachka, A. Farnsworth, A. A. Dhondt, and D. Sheldon (2016b). The implications of mid-latitude climate extremes for North American migratory bird populations. Ecosphere 7:e01261.
La Sorte, F. A., W. M. Hochachka, A. Farnsworth, D. Sheldon, D. Fink, J. Geevarghese, K. Winner, B. M. Van Doren, and S. Kelling (2015b). Migration timing and its determinants for nocturnal migratory birds during autumn migration. Journal of Animal Ecology 84:1202–1212.
La Sorte, F. A., W. M. Hochachka, A. Farnsworth, D. Sheldon, B. M. Van Doren, D. Fink, and S. Kelling (2015c). Seasonal changes in the altitudinal distribution of nocturnally migrating birds during autumn migration. Royal Society Open Science 2: 150347.
LaZerte, S. (2017). feedr: R package for working with data collected from RFID feeder visits. R package version 0.9.0. https://github.com/animalnexus/feedr
Leighton, G. R. M., P. S. Hugo, A. Roulin, and A. Amar (2016). Just Google it: Assessing the use of Google Images to describe geographical variation in visible traits of organisms. Methods in Ecology and Evolution 7:1060–1070.
Lellouch, L., S. Pavoine, F. Jiguet, H. Glotin, and J. Sueur (2014). Monitoring temporal change of bird communities with dissimilarity acoustic indices. Methods in Ecology and Evolution 5:495–505.
Lepczyk, C. A., C. H. Flather, V. C. Radeloff, A. M. Pidgeon, R. B. Hammer, and J. G. Liu (2008). Human impacts on regional avian diversity and abundance. Conservation Biology 22:405–416.
Link, W. A., J. R. Sauer, and D. K. Niven (2017). Model selection for the North American Breeding Bird Survey: A comparison of methods. The Condor: Ornithological Applications 119:546–556.
Lopez-Lopez, P. (2016). Individual-based tracking systems in ornithology: Welcome to the era of big data. Ardeola 63:103–136.
Mason, S. C., G. Palmer, R. Fox, S. Gillings, J. K. Hill, C. D. Thomas, and T. H. Oliver (2015). Geographical range margins of many taxonomic groups continue to shift polewards. Biological Journal of the Linnean Society 115:586–597.
Mayer-Schönberger, V., and K. Cukier (2013). Big Data: A Revolution That Will Transform How We Live, Work, and Think. Houghton Mifflin Harcourt, New York, NY, USA.
Mayor, S. J., R. P. Guralnick, M. W. Tingley, J. Otegui, J. C. Withey, S. C. Elmendorf, M. E. Andrew, S. Leyk, I. S. Pearse, and D. C. Schneider (2017). Increasing phenological asynchrony between spring green-up and arrival of migratory birds. Scientific Reports 7:art.1902.
McCain, C. M., and R. K. Colwell (2011). Assessing the threat to montane biodiversity from discordant shifts in temperature and precipitation in a changing climate. Ecology Letters 14: 1236–1245.
McLaren, J. D., J. J. Buler, T. Schreckengost, J. A. Smolinsky, M. Boone, E. E. van Loon, D. K. Dawson, and E. L. Walters (2018). Artificial light at night confounds broad-scale habitat use by migrating birds. Ecology Letters 21:356–364.
Michener, W. K., and M. B. Jones (2012). Ecoinformatics: Supporting ecology as a data-intensive science. Trends in Ecology & Evolution 27:85–93.
Morrison, C. A., R. A. Robinson, S. J. Butler, J. A. Clark, and J. A. Gill (2016). Demographic drivers of decline and recovery in an Afro-Palaearctic migratory bird population. Proceedings of the Royal Society B 283:20161387.
Nichols, J. D., J. E. Hines, J. R. Sauer, F. W. Fallon, J. E. Fallon, and P. J. Heglund (2000). A double-observer approach for estimating detection probability and abundance from point counts. The Auk 117:393–408.
Pacifici, K., B. J. Reich, D. A. W. Miller, B. Gardner, G. Stauffer, S. Singh, A. McKerrow, and J. A. Collazo (2017). Integrating multiple data sources in species distribution modeling: A framework for data fusion*. Ecology 98:840–850.
Peters, D. P. C., K. M. Havstad, J. Cushing, C. Tweedie, O. Fuentes, and N. Villanueva-Rosales (2014). Harnessing the power of big data: Infusing the scientific method with machine learning to transform ecology. Ecosphere 5:art.67.
Pidgeon, A. M., C. H. Flather, V. C. Radeloff, C. A. Lepczyk, N. S. Keuler, E. M. Wood, S. I. Stewart, and R. B. Hammer (2014). Systematic temporal patterns in the relationship between housing development and forest bird biodiversity. Conservation Biology 28:1291–1301.
Porter, J. H., E. Nagy, T. K. Kratz, P. Hanson, S. L. Collins, and P. Arzberger (2009). New eyes on the world: Advanced sensors for ecology. Bioscience 59:385–397.
Robbins, C. S., D. Bystrak, and P. H. Geissler (1986). The Breeding Bird Survey: Its First Fifteen Years, 1965–1979. Resource Publication 157, U.S. Department of the Interior, Fish and Wildlife Service, Washington, DC, USA.
Robbins, C. S., S. Droege, and J. R. Sauer (1989). Monitoring bird populations with Breeding Bird Survey and atlas data. Annales Zoologici Fennici 26:297–304.
Rowhani, P., C. A. Lepczyk, M. A. Linderman, A. M. Pidgeon, V. C. Radeloff, P. D. Culbert, and E. F. Lambin (2008). Variability in energy influences avian distribution patterns across the USA. Ecosystems 11:854–867.
Sæther, B.-E., and Ø. Bakke (2000). Avian life history variation and contribution of demographic traits to the population growth rate. Ecology 81:642–653.

Saracco, J. F., J. A. Royle, D. F. DeSante, and B. Gardner (2010). Modeling spatial variation in avian survival and residency probabilities. Ecology 91:1885–1891.
Saracco, J. F., J. A. Royle, D. F. DeSante, and B. Gardner (2012). Spatial modeling of survival and residency and application to the Monitoring Avian Productivity and Survivorship program. Journal of Ornithology 152:469–476.
Sarkar, I. N. (2009). Biodiversity informatics: The emergence of a field. BMC Bioinformatics 10 (Suppl. 14):art.S1.
Sauer, J. R., K. L. Pardieck, D. J. Ziolkowski, Jr., A. C. Smith, M.-A. R. Hudson, V. Rodriguez, H. Berlanga, D. K. Niven, and W. A. Link (2017). The first 50 years of the North American Breeding Bird Survey. The Condor: Ornithological Applications 119:576–593.
Sauer, J. R., B. G. Peterjohn, and W. A. Link (1994). Observer differences in the North American Breeding Bird Survey. The Auk 111:50–62.
Schimel, D., and M. Keller (2015). Big questions, big science: Meeting the challenges of global ecology. Oecologia 177: 925–934.
Shonfield, J., and E. M. Bayne (2017). Autonomous recording units in avian ecological research: Current use and future applications. Avian Conservation and Ecology 12:art.14.
Sillett, T. S., and R. T. Holmes (2002). Variation in survivorship of a migratory songbird throughout its annual cycle. Journal of Animal Ecology 71:296–308.
Stephens, P. A., L. R. Mason, R. E. Green, R. D. Gregory, J. R. Sauer, J. Alison, A. Aunins, L. Brotons, S. H. M. Butchart, T. Campedelli, T. Chodkiewicz, et al. (2016). Consistent response of bird populations to climate change on two continents. Science 352:84–87.
Stowell, D., M. Wood, Y. Stylianou, and H. Glotin (2016). Bird detection in audio: A survey and a challenge. In Proceedings of MLSP 2016: 2016 IEEE International Workshop on Machine Learning for Signal Processing, September 13–16, Vietri sul Mare, Salerno, Italy (F. A. N. Palmieri, A. Uncini, K. Diamantaras, and J. Larsen, Editors). Institute of Electrical and Electronics Engineers, New York, NY, USA. pp. 1–6.
Strasser, B. J. (2012). Data-driven sciences: From wonder cabinets to electronic databases. Studies in History and Philosophy of Science, Part C 43:85–87.
Strebel, N., M. Kéry, M. Schaub, and H. Schmid (2014). Studying phenology by flexible modelling of seasonal detectability peaks. Methods in Ecology and Evolution 5:483–490.
Strong, C., B. Zuckerberg, J. L. Betancourt, and W. D. Koenig (2015). Climatic dipoles drive two principal modes of North American boreal bird irruption. Proceedings of the National Academy of Sciences USA 112:E2795–E2802.
Stutchbury, B. J. M., S. A. Tarof, T. Done, E. Gow, P. M. Kramer, J. Tautin, J. W. Fox, and V. Afanasyev (2009). Tracking long-distance songbird migration by using geolocators. Science 323:896.
Sullivan, B. L., J. L. Aycrigg, J. H. Barry, R. E. Bonney, N. Bruns, C. B. Cooper, T. Damoulas, A. A. Dhondt, T. Dietterich, A. Farnsworth, D. Fink, et al. (2014). The eBird enterprise: An integrated approach to development and application of citizen science. Biological Conservation 169:31–40.
Taylor, P. D., T. L. Crewe, S. A. Mackenzie, D. Lepage, Y. Aubry, Z. Crysler, G. Finney, C. M. Francis, C. G. Guglielmo, D. J. Hamilton, R. L. Holberton, et al. (2017). The Motus Wildlife Tracking System: A collaborative research network to enhance the understanding of wildlife movement. Avian Conservation and Ecology 12:art.8.
Tien, J. M. (2013). Big data: Unleashing information. Journal of Systems Science and Systems Engineering 22:127–151.
Tingley, M. W., M. S. Koo, C. Moritz, A. C. Rush, and S. R. Beissinger (2012). The push and pull of climate change causes heterogeneous shifts in avian elevational ranges. Global Change Biology 18:3279–3290.
Vallejo, E. E., and C. E. Taylor (2009). Adaptive sensor arrays for acoustic monitoring of bird behavior and diversity: Preliminary results on source identification using support vector machines. Artificial Life and Robotics 14:485–488.
VanDerWal, J., H. T. Murphy, A. S. Kutt, G. C. Perkins, B. L. Bateman, J. J. Perry, and A. E. Reside (2013). Focus on poleward shifts in species' distribution underestimates the fingerprint of climate change. Nature Climate Change 3:239–243.
Van Doren, B. M., K. G. Horton, A. M. Dokter, H. Klinck, S. B. Elbin, and A. Farnsworth (2017). High-intensity urban light installation dramatically alters nocturnal bird migration. Proceedings of the National Academy of Sciences USA 114:11175–11180.
Veech, J. A., K. L. Pardieck, and D. J. Ziolkowski, Jr. (2017). How well do route survey areas represent landscapes at larger spatial extents? An analysis of land cover composition along Breeding Bird Survey routes. The Condor: Ornithological Applications 119:607–615.
Watson, M. J., D. R. Wilson, and D. J. Mennill (2016). Anthropogenic light is associated with increased vocal activity by nocturnally migrating birds. The Condor: Ornithological Applications 118:338–344.
White, E. P., and A. H. Hurlbert (2010). The combined influence of the local environment and regional enrichment on bird species richness. The American Naturalist 175:E35–E43.
Wilmers, C. C., B. Nickel, C. M. Bryce, J. A. Smith, R. E. Wheat, and V. Yovovich (2015). The golden age of bio-logging: How animal-borne sensors are advancing the frontiers of ecology. Ecology 96:1741–1753.
Wimmer, J., M. Towsey, P. Roe, and I. Williamson (2013). Sampling environmental acoustic recordings to determine bird species richness. Ecological Applications 23:1419–1428.
Witten, I. H., E. Frank, M. A. Hall, and C. J. Pal (2016). Data Mining: Practical Machine Learning Tools and Techniques, fourth edition. Morgan Kaufmann/Elsevier, Cambridge, MA, USA.
Zhao, Z., S.-h. Zhang, Z.-y. Xu, K. Bellisario, N.-h. Dai, H. Omrani, and B. C. Pijanowski (2017). Automated bird acoustic event detection and robust species classification. Ecological Informatics 39:99–108.
Zuckerberg, B., D. Fink, F. A. La Sorte, W. M. Hochachka, and S. Kelling (2016). Novel seasonal land cover associations for eastern North American forest birds identified through dynamic species distribution modelling. Diversity and Distributions 22:717–730.