Simulating avian body mass measurements using `birdsize`

Introduction

Different currencies of measurement - e.g. total number of individuals, total biomass, or total metabolic flux or energy use - provide linked, but qualitatively very different, perspectives on the structure and function of ecological systems (White et al. (2007)). The study of the interrelated dynamics of size structure, species composition, individual abundance, and biomass and energy use is well-established for systems for which data on both individuals' body sizes and individual organismal abundance are widely available, including aquatic systems, terrestrial forest systems, and, to a lesser extent, small mammal systems (Kerr and Dickie (2001), White et al. (2007)). Work in these systems has yielded important insight into - for example - how ecological degradation can manifest in the relationship between total abundance and total biomass (Warwick and Clarke (1994)), or how shifts in community-wide mean body size can buffer total energy use against apparent changes in total individual abundance (White et al. (2004)). Efforts to generalize these efforts to terrestrial vertebrate systems have been constrained due to the lack of body size measurements for these communities (White et al. (2007), Thibault et al. (2011)). Sampling methodologies for avian communities often rely on visual or auditory point-counts, which provide information about species abundance and diversity but do not directly capture information about body size or energy use.

The `birdsize` R package offers a way around this limitation by estimating individual-level (and, from there, population or community-wide) body size measurements for birds given either species identity or a species' mean and/or standard deviation of body size. Birds exhibit determinate growth, and `birdsize` assumes that intraspecific body size distributions for birds are, to a first approximation, well-described by normal distributions parameterized with a species-specific mean and standard deviation (see also Thibault et al. (2011)). Moreover, there is a strong scaling relationship between a species' mean body size and its standard deviation of body size, meaning that, for species for which the standard deviation is not known, the standard deviation can be estimated from the mean (see also Thibault et al. (2011)). Estimates obtained in this way are, of course, considerably less precise than those that could be obtained through exhaustive field sampling, and may not be appropriate for all use cases. However, given the logistical constraints on field operations of this scale (and the even harsher constraint of time, which prevents us from retroactively taking these measurements for ecological timeseries), `birdsize` makes it possible to conduct macroecological-scale analyses of avian communities that would not otherwise be possible. This approach was first used at scale by Thibault et al. (2011) and subsequently by Diaz and Ernest (2022) (in review). `birdsize` formalizes this method and makes it accessible via a straightforward user interface, in order to facilitate use by other research groups with diverse use cases.

The estimation procedure in `birdsize`

The core functionality of `birdsize` is to generate estimates of individual body size for populations of birds by drawing from a normal distribution parameterized with a species-level mean and standard deviation of body size. It includes built-in values for these parameters for 443 species found in the North American Breeding Bird Survey (Pardieck et al. (2019)), and can accept user-supplied parameter values for additional species.

For the 443 species included with `birdsize`, mean and standard deviation values were manually obtained from the CRC Handbook of Avian Body Masses (Dunning (2008)). These species are listed in the data frame `birdsize::known_species`. For records in Dunning (2008) with mean, but no standard deviation, reported, the standard deviation is estimated via a scaling relationship between the mean and standard deviation of body mass (see also Thibault et al. (2011)). Specifically, a linear model of the form `log(variance(body_size)) ~ log(mean(body_size))` has a model R^2 of 0.89, and produces the scaling relationship of `variance(body_size) = 0.0047(body_size) ^ 2.01`. This scaling relationship is used to generate estimated standard deviations for records without standard deviation recorded, affecting 353 of 928 raw records. Finally, many species in Dunning (2008) have multiple records from different time periods, locations, and subspecies. In these instances, parameter values are averaged across records to obtain a single species-wide value.

A user may also manually supply parameter values, in order to generate estimates for species not included in `birdsize::knownspecies`, or to use different parameter values than those included with `birdsize`. This may be of particular interest for users wishing to explore questions related to (for example) intraspecific variation in body size across different populations of the same species affects, or extending to species not common found North America. In this case, if both mean and standard deviation are supplied, they will be used, and if only the mean is provided, the standard deviation is estimated via the scaling relationship explained above.

Population and community-wide summaries

Integration with the Breeding Bird Survey

The methodology in `birdsize` was first developed and applied to the North American Breeding Bird Survey, and `birdsize` is built to naturally accommodate Breeding Bird Survey data obtained from ScienceBase (Pardieck et al. (2019)) or tools such as the Data Retriever (Senyondo et al. (2017)). There is no actual data from the Breeding Bird Survey included in the `birdsize` package, and users are encouraged to access the most up-to-date data from the creators directly. To facilitate this, the `bbs-data` and demonstration vignettes illustrate how to access these data and use them with `birdsize`, and the example data tables in `birdsize` (i.e. `demo_route_raw` and `demo_route_clean`) contain synthetic data matching the format of the Breeding Bird Survey.

However, `birdsize` is not constrained to work *only* with Breeding Bird Survey data. It accepts any dataset, real or synthetic, that includes population sizes and species identity and/or body size parameters (see above); see Use case #3, below.

Use case 1: Simulation over the Breeding Bird Survey timeseries

Use case 2: Simulating temporal shifts in body size based on the Breeding Bird Survey

Use case 3: Simulating imaginary birds

References

Diaz, R. M., and S. K. M. Ernest. 2022, November. Temporal changes in the individual size distribution decouple long-term trends in abundance, biomass, and energy use of North American breeding bird communities. bioRxiv.

Dunning, J. B. 2008. CRC handbook of avian body masses. CRC handbook of avian body masses. 2nd ed. CRC Press, Boca Raton.

Kerr, S. R., and L. M. Dickie. 2001. The Biomass Spectrum: A Predator-Prey Theory of Aquatic Production. Pages 352 Pages. Columbia University Press.

Pardieck, K. L., D. J. Ziolkowski, M. Lutmerding, V. Aponte, and M.-A. Hudson. 2019. North American Breeding Bird Survey Dataset 1966 - 2018, version 2018.0. U.S. Geological Survey.

Senyondo, H., B. D. Morris, A. Goel, A. Zhang, A. Narasimha, S. Negi, D. J. Harris, D. G. Digges, K. Kumar, A. Jain, K. Pal, K. Amipara, and E. P. White. 2017. Retriever: Data Retrieval Tool. Journal of Open Source Software 2:451.

Thibault, K. M., E. P. White, A. H. Hurlbert, and S. K. M. Ernest. 2011. Multimodality in the individual size distributions of bird communities. Global Ecology and Biogeography 20:145–153.

Warwick, R. M., and K. R. Clarke. 1994. Relearning the ABC: Taxonomic changes and abundance/biomass relationships in disturbed benthic communities. Marine Biology 118:739–744.

White, E. P., S. K. M. Ernest, A. J. Kerkhoff, and B. J. Enquist. 2007. Relationships between body size and abundance in ecology. Trends in Ecology & Evolution 22:323–330.

White, E. P., S. K. M. Ernest, and K. M. Thibault. 2004. Trade-offs in Community Properties through Time in a Desert Rodent Community. The American Naturalist 164:670–676.