Best practices for making reliable inferences from citizen science data: case study using eBird to estimate species distributions

A Johnston, WM Hochachka, ME Strimas-Mackey, V Ruiz Gutierrez, OJ Robinson, ET Miller, T Auer, ST Kelling, D Fink

SUPPORTING INFORMATION APPENDIX A1

Template for describing eBird data used in analyses

The following lists the minimum information that should be included when describing eBird data selected and subsequent data processing.

Data selected

- Version and date of the eBird dataset:
- Spatial extent: latitude and longitude limits, or country/state geographic filters
- Temporal extent: years included and date filters within years
- Species: species included and any bespoke sub-species or supra-species selection

Data processing

- Zero-filling (yes or no)
- Treatment of X (presence-only observations with no count)
- Treatment of group checklists
- Any further filtering criteria (e.g. protocol, complete checklists, only checklists with certain information)

Example 1

We used eBird data for dunlin from the eBird DataBase (EBD) version released in August 2018 (Sullivan et al., 2009; eBird, 2018). We filtered the checklists to those from Jan 1 2010 to Dec 31 2017 and within the state of California using the *auk* R package (Strimas-Mackey, Miller, and Hochachka 2017). We selected 'complete checklists' with durations of up to 3 hours and up to 10 observers. We selected checklists with the 'stationary' protocol or the 'traveling' protocol, with a maximum distance travelled of 5km. We removed any checklists without a complete suite of effort information (time observations started, duration, protocol, distance travelled, number of observers). We also removed any checklists with presence of dunlin, but no counts (X). We used absence of dunlin on a complete checklist as an indication of a zero-count and zero-filled the data.

Example 2

We used eBird data from the eBird DataBase (EBD) version released in August 2018 (Sullivan et al., 2009; eBird, 2018). We filtered the checklists to those from Jan 1 2016 to Dec 31 2017 and within Mexico using the *auk* R package (Strimas-Mackey, Miller, and Hochachka 2017). We selected 'complete checklists' from all protocols and with a maximum distance travelled of 10km. We removed any checklists without a complete suite of effort information (time observations started, duration, protocol, distance travelled, number of observers). We calculated the number of species recorded on each checklist, including 'detections' as species counts or species presence-only (X).

References

eBird. (2018). eBird: An online database of bird distribution and abundance. Retrieved 20 October 2018, from www.ebird.org

Strimas-Mackey, M., Miller, E., & Hochachka, W. (2017). auk: eBird Data Extraction and Processing with AWK. *R Package Version*.

Sullivan, B. L., Wood, C. L., Illiff, M. J., Bonney, R. E., Fink, D., & Kelling, S. (2009). eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, *142*(10), 2282–2292.