

## Identifying engaging bird species and traits with community science data

Birders and amateur naturalists engage with different bird species at different rates. Quantifying these species-specific observer biases can inform the impact and efficacy of conservation outreach and the scientific use of community collected biodiversity data. Species that are thought to be “charismatic” or more engaging are often prioritized in conservation, and previous researchers have used sociological experiments and digital records to estimate charisma indirectly. We developed methods to estimate species engagement directly from citizen science data as a relative reporting rate between iNaturalist and eBird data across the continental USA. We assume eBird complete checklists represent the rates at which species-agnostic observers detect each species, whereas the more opportunistic iNaturalist data represent selective engagements based on user preference and logistical factors such as site selection and observer skill. We define an “overreporting index” for each species as the geographic median log difference between iNaturalist and eBird reporting rates. The spatiotemporal scale of these data raise several challenges. To accommodate spatial correlation, we estimate spatial surfaces of user engagement as generalized additive models (GAMs) for each of 472 species of birds across the contiguous United States in each dataset. To accommodate other sources of heterogeneity, we use quasibinomial responses. To propagate uncertainty from the GAMs to uncertainty in the overreporting index for subsequent analysis, we quantify first-stage uncertainty with bootstrapping, then use a mixed effects meta-analysis across species. In the meta-analysis, we ask whether iNaturalist users engage more with larger, more colorful, and rarer birds relative to eBird users. We find evidence that, across 472 modeled species, 52 species are significantly overreported and 158 are significantly underreported, indicating a wide variety of species-specific effects. We also find that body mass, color contrast, and range size all predict overreporting in the opportunistic dataset. The two-stage methodology presented for propagating uncertainty from computationally demanding species-specific models to across-species meta-analysis can be customized spatiotemporally and incorporated into other kinds of ecological statistical models. Understanding which bird species are highly engaging can aid conservation and outreach efforts.