

Modelling species abundance across large spatial extents: opportunities and challenges

A. Johnston ^{a1}, D. Fink ^a, M.D. Reynolds ^b, N.E. Bruns ^a, W.M. Hochachka ^a, F.A. La Sorte ^a
and S. Kelling ^a

^a Cornell Lab of Ornithology
Cornell University
Ithaca, NY 14850, USA
¹aj327@cornell.edu

^b The Nature Conservancy
201 Mission Street, 4th Floor
San Francisco, CA 94105, USA

Keywords: abundance; citizen science; species distribution models.

Abstract: Information about the relative abundance of species can reveal much about the patterns, processes and dynamics which may be underlying natural systems. To date, most species distribution modelling has been concerned with the presence or occurrence of species in space and time. Abundance data are often reduced to presence-absence data to estimate species distributions, resulting in loss of considerable information. When considering abundance data collected over wide spatial extents, a number of statistical challenges are often present, including zero-inflation, overdispersion, and non-stationarity. Species' behavioural characteristics such as seasonal aggregations can also lead to statistical problems. Here we discuss and present various parametric and machine learning models adapted for these challenges. We use data from eBird a large-scale citizen science project, which contains counts of bird species alongside location and search effort information. We examine the performance of the models, which also account for the inherent variability in effort that occurs with citizen science data. We consider suitable model validation metrics for abundance data. We show that modelling abundance reveals patterns and processes that are not evident when considering presence-absence data, which enables ecological research that is more detailed and mechanistic, and supports better informed conservation and management applications.