Running head: OCCUPANCY MODELS FOR HISTORICAL DATA

Utilizing occupancy-detection models with museum specimen data: promise and pitfalls

Laura Melissa Guzman[1], Rassim Khelifa[2], Leithen K. M'Gonigle[1], Leslie Ries[3], Vaughn Shirey[3]

Names currently alphabetical

1. Department of Biological Sciences
   Simon Fraser University
   8888 University Drive, Burnaby, BC, Canada,V5A 1S6

2. Department of Zoology
   University of British Columbia
   6270 University Blvd, Vancouver, BC, Canada,V6T 1Z4

3. Department of Biology
   Georgetown University
   37th and O Streets NW, Washington, DC, United States, 20057

Abstract

Historical museum records provide potentially useful data for identifying drivers of temporal trends in species occupancy, however, because these records were not explicitly collected for this purpose, methodological developments are needed in order to enable robust inferences. Occupancy-detection models, a relatively new and powerful suite of methods, are potentially useful here, because these models allow us to account for changes in collection effort through space and time. Applying such occupancy-detection models to historical museum records is not a straightforward process, as these models have strict data requirements that museum data usually do not meet. Failing to account for these requirements can lead to biased results (e.g., Guzman et al., 2021). Here we present a methodological road-map for using occupancy models to analyze historical museum records. We use simulated data-sets to identify how and when modelling decisions and patterns in data can bias inferences. We focus primarily on the consequences of contrasting methodological approaches for dealing with species ranges and non-detections in both space and time (Figure 1). Finally, we include a case study using butterflies in California, in order to compare the inferences produced from museum records vs. those derived from rigorous long-term monitoring schemes. Our work also includes computational optimizations that will faciliate analyses of large data-sets.

Keywords: global change, occupancy, museum specimens, hierarchical model

Please briefly explain why the proposed paper is in scope for the Special Feature and relevant to the Journal:

The use of occurrence data from museum collections is becoming more common in ecological analyses over the last several years. Importantly, however, improper treatment of these data can lead to misleading inferences (Guzman et al., 2021; Larsen and Shirey, 2021). Our proposed paper will survey the methodological challenges one must overcome in order to successfully apply occupancy models to museum data. Specifically, we will focus on consequences of: inferring past episodes of site visitation in time and space, the influence of heterogeneity in site-visitation history, and the influence of modelling spatial scales that include regions outside some species' geographic ranges. Using simulated data (which enables us to evaluate model performance against known trends), we will generate a checklist of data and model considerations, highlighting potential consequences of violations.

Our paper will fill an important gap in the toolbox of methods aimed at identifying drivers of change in species distributions. There is an urgent need to identify and mitigate drivers of species range collapse and, therefore, it is crucial that we have powerful analytical tools with corresponding guidelines for best practices. Our paper will provide such a tool and is, therefore, compatible with the proposed Special issue "Leveraging Natural History Collections to Understand the Impacts of Global Change" in Methods in Ecology and Evolution. This work will be of wide interest to researchers in both ecology and evolution.

References

Guzman, L. M., S. A. Johnson, A. O. Mooers, and L. K. M'Gonigle, 2021. Using historical data to estimate bumble bee occurrence: Variable trends across species provide little support for community-level declines. Biological Conservation 257:109141.

Larsen, E. A. and V. Shirey, 2021. Method matters: pitfalls in analysing phenology from occurrence records. Ecol. Lett. P. Early View.

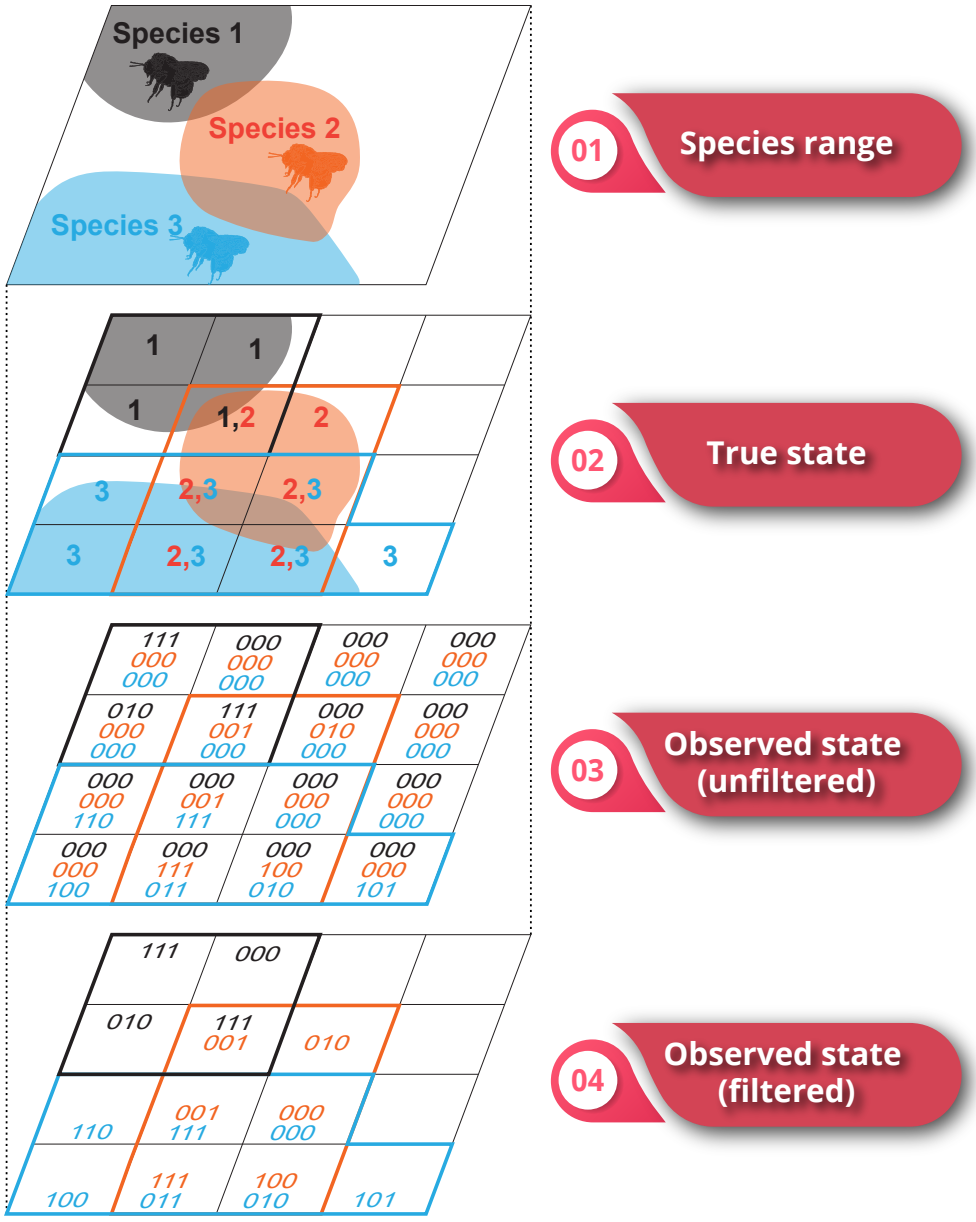

Figure 1: Workflow used to analyze multi-species occupancy models. The observation process is constructed based on detection-nondetection data of multiple taxa. The detection of a species is recorded when it is observed at a given site in a given visit (visit-site). Since most surveys do not record the absence of species, non-detections have to be inferred. In our approach, the generated absences are also restricted to the species range, which not only speeds up the computation of the model but also makes the estimates of detection and occupancy probabilities more reliable.