# Introduction to citizen science data

Diana Bowler

Gfoe workshop,

29th August 2021

# Outline

- diversity of citizen science data
- bias associated with each (site selection, reporting bias, detection issues …)
- simple trend models

# Diversity of citizen science

## Structured data

- Standardized sampling protocol
- Site-selection – sometimes stratified random, often not

## Semi-structured data

- No standardized sampling protocol
- Site selection - free
- Metadata associated with data informs on survey methods

## Unstructured data

- No standardized sampling protocol
- Site selection – free
- Little metadata

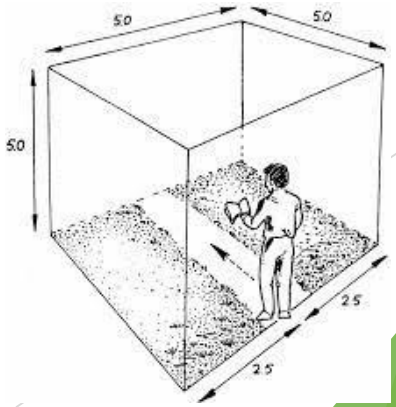
# Structured citizen science

- Analysis is (relatively) easy!!

- Still can be site selection biases and missing data issues (see Bled et al. 2013; Sauer et al. 2020; bbBayes)

The North American Breeding Bird Survey


tagfalter-monitoring.de



# Semi-structured and structured data

- Examples:

iNaturalist

eBird

FLORA incognita

naturgucker.de

GBIF
Global Biodiversity Information Facility

Observation.org

# Semi-structured and structured data

- Analysis more difficult!
- People are not coordinated in their efforts
- Variation among people in how they collect data

# How do people vary in data collection?

**Table 1.** Traits of recorders that could be influential in describing different recorder 'profiles' or 'syndromes'; a range of potential profiles have been identified

| Trait | Relevance to information content |
|---|---|
| Complete lists? | An indication of the typical effort per survey |
| Coverage of 'rare' species | Predilection for reporting unusual sightings |
| Coverage of difficult species | Taxonomic expertise |
| Length of activity of reporting | Temporal footprint |
| Frequency of recording | Productivity and consistency |
| Spatial variation in recording | Spatial footprint of the data |
| Variation in recording across taxa | Consistency of recording across taxa (taxonomic specialist vers |

Isaac and Pocock, 2015

# Semi-structured and structured data

- Analysis more difficult!
- Need to consider observation/sampling processes:
  - People are not coordinated in their efforts
  - Variation among people in how they collect data
- Need to model these observation/sampling processes
  - Estimating detection probability
  - ‘Imperfect detection’

# Concept of imperfect detection

- When we do a wildlife survey, we (almost) never see all individuals of a species present.

# Is imperfect detection always a problem?

- No – we don't need to worry about imperfect detection when:
  - We can assume detection probability don't change over time or space

  and
  - we are only interested in species' relative occurrences/abundances and not absolute values.
- We make this assumption most of the time when we analyze structured citizen science data

# When can imperfect detection be a problem?

- We are comparing among habitats

- We are comparing among species

- We are comparing among surveys collected with different methods or durations

# Imperfect reporting too!!

- Most unstructured citizen science observations are just of one species … probably more species were seen!!!
  - Presence-only data

- For semi-structured citizen science, we might have metadata on whether an observation comes from a ‘complete checklist’
  - Presence-absence data

Birds present

Birds observed

Birds recorded

# Hierarchical models for citizen science data

- Models to ask questions about:
- How large is the population or distribution of my species?
- What factors explain where my species lives?
- Is my species declining or increasing over time?

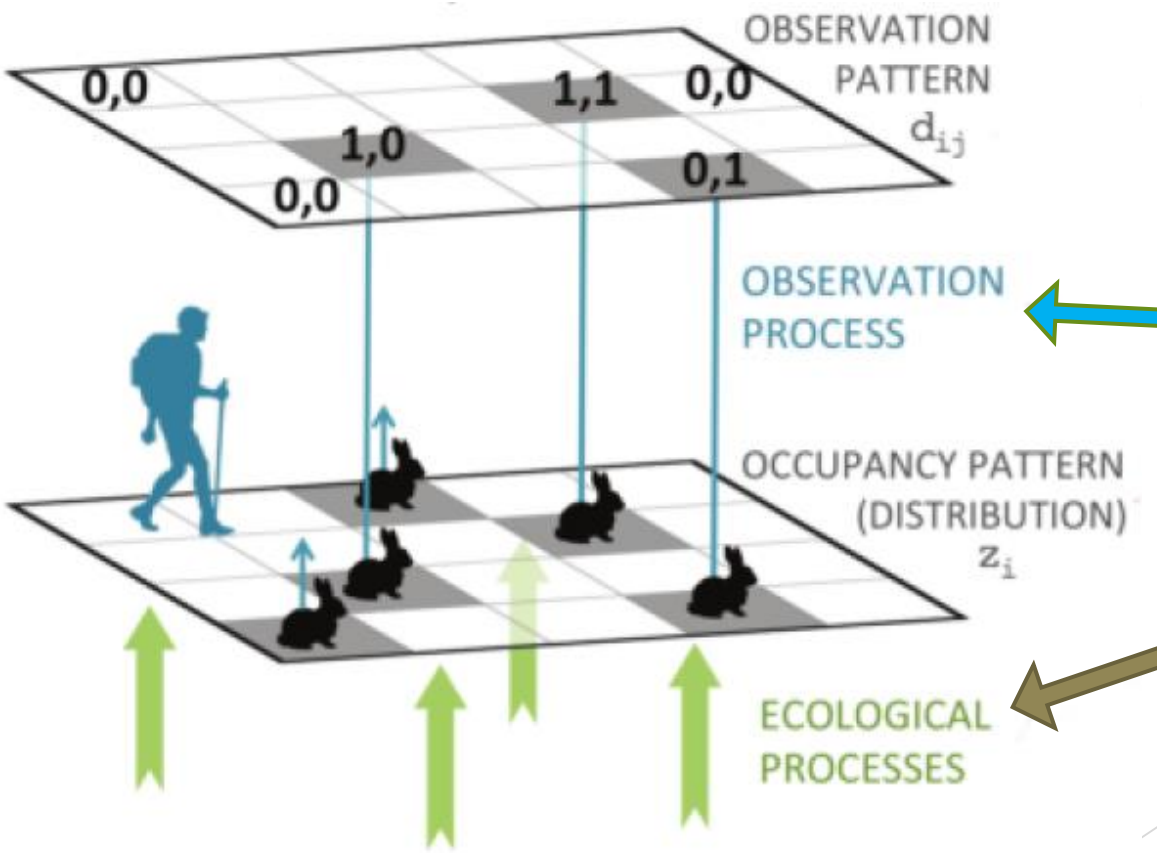
# Different types of related models

- Occupancy Models
- N-Mixture Models
- Distance sampling Models

- All these models assume data is generated by:
  - ecological processes affecting where the species is
  - observation (or sampling) processes affecting where the species is detected by my survey method

- We have separate models for each process

# Hierarchical models to account for observation/sampling processes

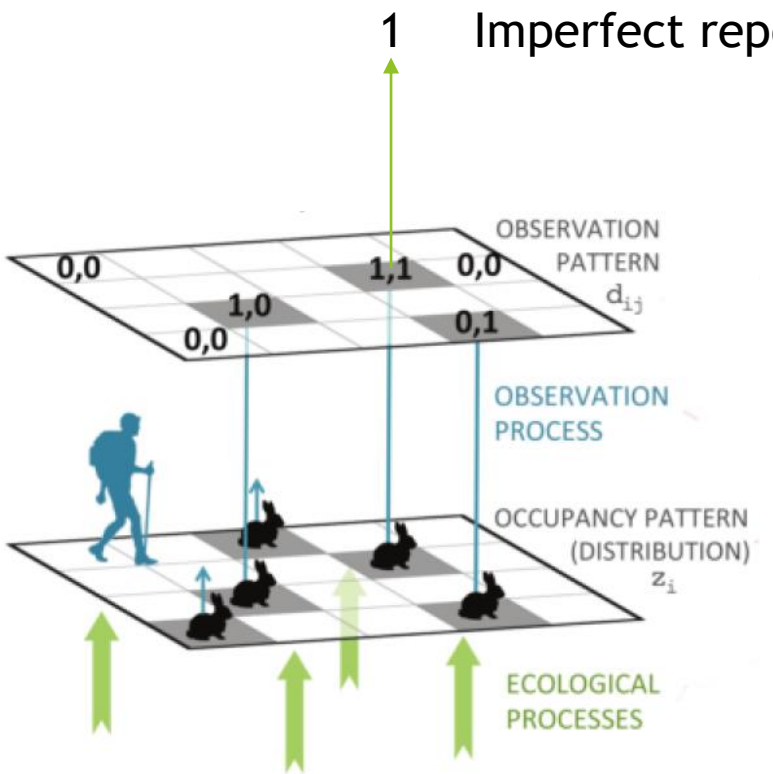



Usually modelled separately

Image by: Res Altwegg

# Imperfect detection in CS includes imperfect reporting

- In unstructured citizen science, people might see more species than they report



# Common detection covariates

| Type | Direction |
|---|---|
| Sampling effort | Species are more detectable with longer survey duration |
| Checklist length | Single list (opportunistic) or longer (complete checklist?) |
| Date of year | Dependent on species' phenology (e.g., flight period of butterflies) |
| Observer | More experienced observers might be able to detect species more often |
| Forest cover | Species harder to see in forest |
| Climate | Some species (e.g., many insects) are amore active on warm, sunny, rain-free days |
| …… | |

Depends on survey method: visual, acoustic, DNA-based etc..

# Simple trend analysis

- Response: Count (Poisson or negative binomial) or Occurrence (binomial)
- Predictors: Site and Year (as fixed or random effects)

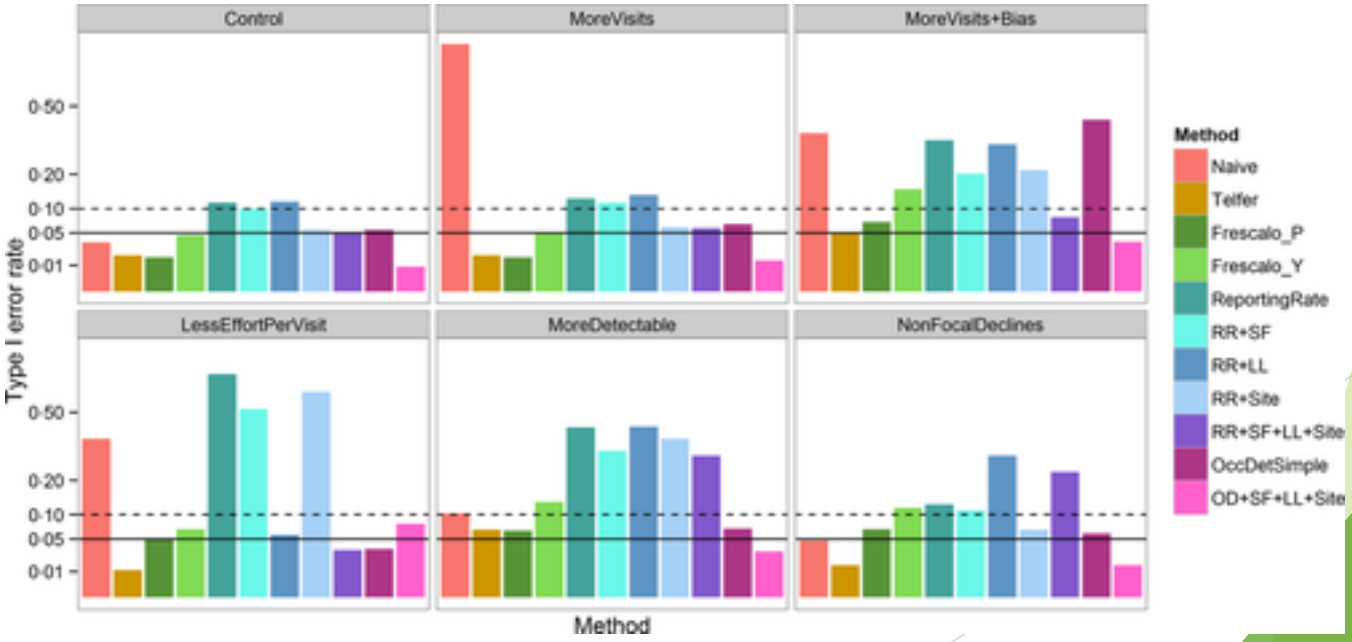
# Useful resources

Methods in Ecology and Evolution | BRITISH ECOLOGICAL SOCIETY

Research Article | Open Access

## Statistics for citizen science: extracting signals of change from noisy ecological data

Nick J. B. Isaac, Arco J. van Strien, Tom A. August, Marnix P. de Zeeuw, David B. Roy



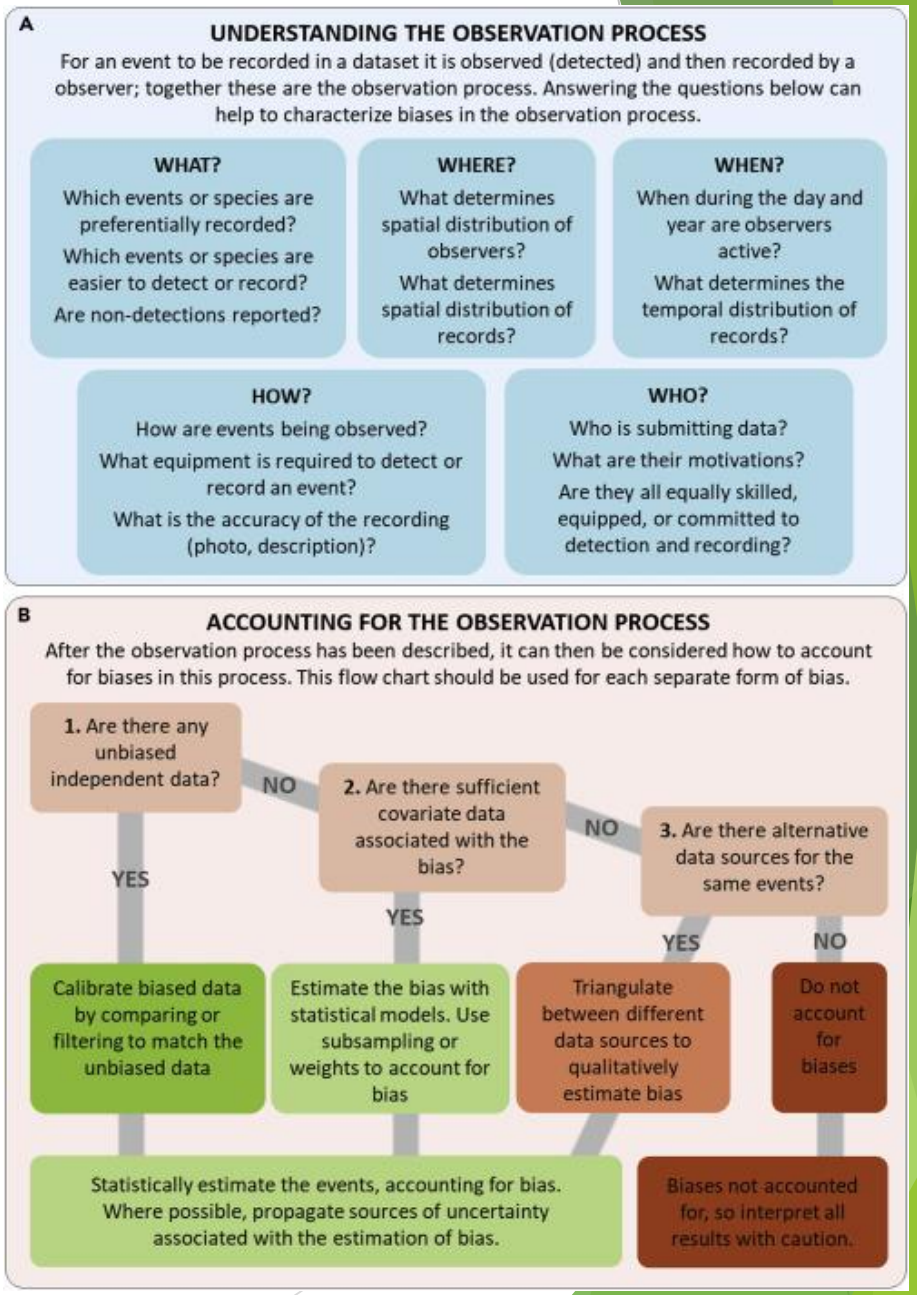
# Useful resources

**One Earth**
Volume 2, Issue 5, 22 May 2020, Pages 455-465

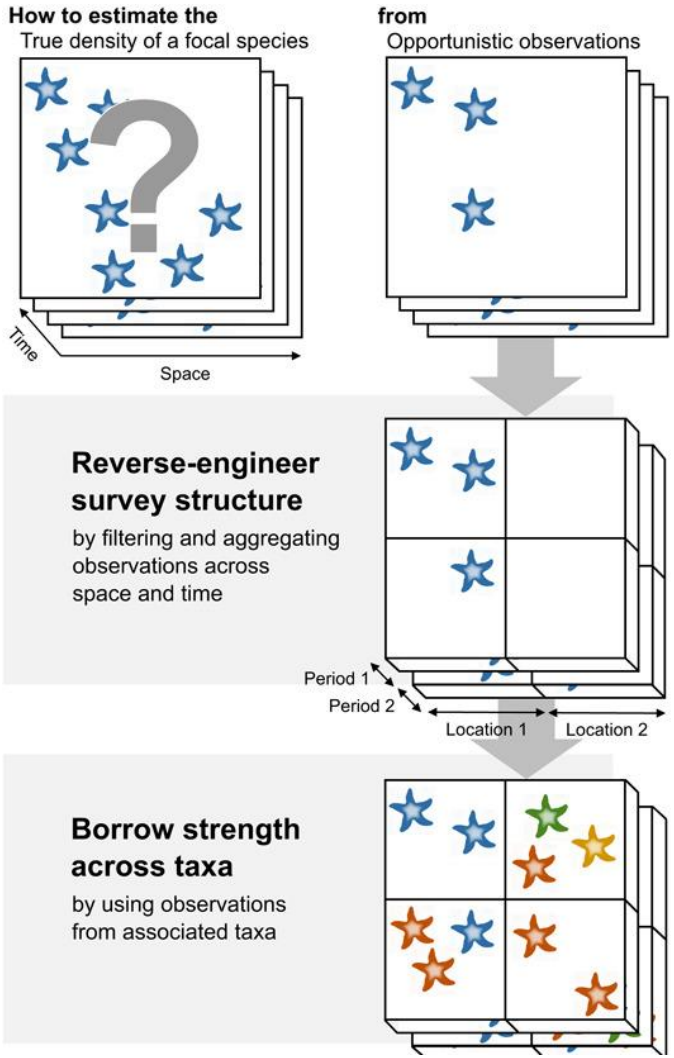
Perspective

## Making Messy Data Work for Conservation

A.D.M. Dobson [1], E.J. Milner-Gulland [2], Nicholas J. Aebischer [3], Colin M. Beale [4], Robert Brozovic [5], Peter Coals [6], Rob Critchlow [4], Anthony Dancer [7], Michelle Greve [8], Amy Hinsley [6], Harriet Ibbett [9], Alison Johnston [10], Timothy Kuiper [2], Steven Le Comber [11, 20], Simon P. Mahood [12, 13], Jennifer F. Moore [14], Erlend B. Nilsen [15], Michael J.O. Pocock [16] ... Aidan Keane [1]



# Useful resources




OIKOS
ADVANCING ECOLOGY
Forum | Full Access

**Deriving indicators of biodiversity change from unstructured community-contributed data**

Giovanni Rapacciuolo, Alison Young, Rebecca Johnson

First published: 04 June 2021 | https://doi.org/10.1111/oik.08215 | Citations: 1