Introduction and background

Point count data analysis workshop 2025

Péter Sólymos

2025-11-08

Housekeeping

Online etiquette

- Camera on optional
- Go for a break as needed (2 breaks/session)
- Stretch

Questions

- Live
- In the chat
- Or https://hackmd.io/@psolymos/APCD01

Course materials

- Short URL: a8m.io/pc
- `git clone https://github.com/psolymos/point-count-data-analysis.git`
- Download zip file from GitHub

Teaching format

- Lectures: me talking, slides, Q&A
- Hands-on practices: follow along or just watch, R/Quarto code
- Interactive exploration: R/Quarto, Shiny apps

About the instructor

- **Péter Sólymos**
- Senior Data Scientist @ E Source
- Adjunct Professor @ Univ. Alberta
- Living in Edmonton, Alberta, Canada, Treaty 6 territory
- R programmer (vegan, detect, ResourceSelection, pbapply)
- peter.solymos.org

What you will learn

- Understand basic statistical concepts related to detection error
- Work with field collected data and data from automated recording units (ARU)
- Know packages such as unmarked, detect, bSims
- Critically evaluate modelling options and assumptions using simulations
- Fit N-mixture, distance sampling, and time-removal models to data

In short

1. How to analyze your point count data when
2. How to violate assumptions and get away with it

Outline

- Day 1: Laying the groundwork [4 hrs]
- Day 2: Understanding mechanisms [4 hrs]
- Day 3: Advanced topics [4 hrs]

Day 1: Laying the groundwork

- Introduction and background [30 min]
- Review of field sampling techniques [30 min]
- Overview of regression techniques [90 min]
- Naïve estimates of occupancy and abundance [30 min]
- Multiple visits and N-mixture models [30 min]

Day 2: Understanding mechanisms

- Introduction to agent-based simulations [30 min]
- Bird behavior [30 min]
- Time-removal models [30 min]
- Observation process [30 min]
- Distance sampling [30 min]
- Combining removal and distance sampling (QPAD) [30 min]
- Conducting simulation experiments [30 min]

Day 3: Advanced topics

- Single visit-based approaches (N-mixture and SQPAD) [60 min]
- Dealing with roadside and other biases [60 min]
- Analyzing data from recording units [30 min]
- Multi-species models and using species traits and phylogeny [30 min]
- Closing remarks & Q&A [30 min]

Prerequisites

A basic understanding of statistical, mathematical and physical concepts. Specifically, generalised linear regression models, including mixed models; basic knowledge of calculus.

Familiarity with R, ability to import/export data, manipulate data frames, fit basic statistical models (up to GLM) and generate simple exploratory and diagnostic plots.

See Day 0 on GitHub: links to free resources.

Setup instructions

- Install R
- Install RStudio/VS Code/Positron/Emacs, etc.

Install required packages with the `install.R` script:

```r
source(paste0("https://raw.githubusercontent.com/",
    "psolymos/point-count-data-analysis/refs/heads/main/day-00/install.R"))
```

See Day 0 on GitHub.

Next

Review of field sampling techniques

Field sampling techniques

Apples and oranges

"A comparison of apples and oranges occurs when two items or groups of items are compared that cannot be practically compared." [*Wikipedia*]

What we measure

How we measure things can have big impact on our results.

- You might say: I saw 5 magpies (walking down the street),
- I might say: I only saw one (sitting on my deck)

Apples to apples

Effort:

- *area* of the physical space searched
- amount of *time* spent
- number of *individuals* identified

Experience, skill etc.:

- number of years in field work
- eye sight, hearing ability
- microphone sensitivity, damage

The goal is to make our measurements **comparable**.
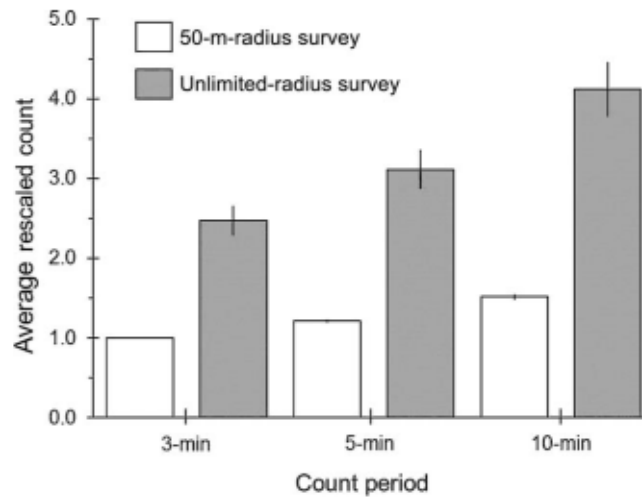
Effects can be significant

Figure 1: Matsuoka et al. 2014

10-min unlimited count ~300% increase over 3-min 50-m count. Average across 54 species of boreal songbirds[1].

What is a point count?

- A trained observer
- records all the birds
- seen and heard
- from a point count station
- for a set period of time
- within a defined distance radius.

Questions we want to answer using point counts

- How many? (Abundance, density, population size)
- Is this location part of the range? (0/1)
- How is abundance changing in space? (Distribution)
- How is abundance changing in time? (Trend)
- What is the effect of a treatment on (relative) abundance?

---

[1]Ralph et al. 1993, Handbook of field methods for monitoring landbirds.

Design based standardization

Have a set of standards/recommendations that people will follow to

- Maximize efficiency in the numbers of birds and species counted
- Minimize extraneous variability in the counts[^2]

But programs started to deviate from standards:

"For example, only 3% of 196,000 point counts conducted during the period 1992–2011 across Alaska and Canada followed the standards recommended for the count period and count radius."[2]
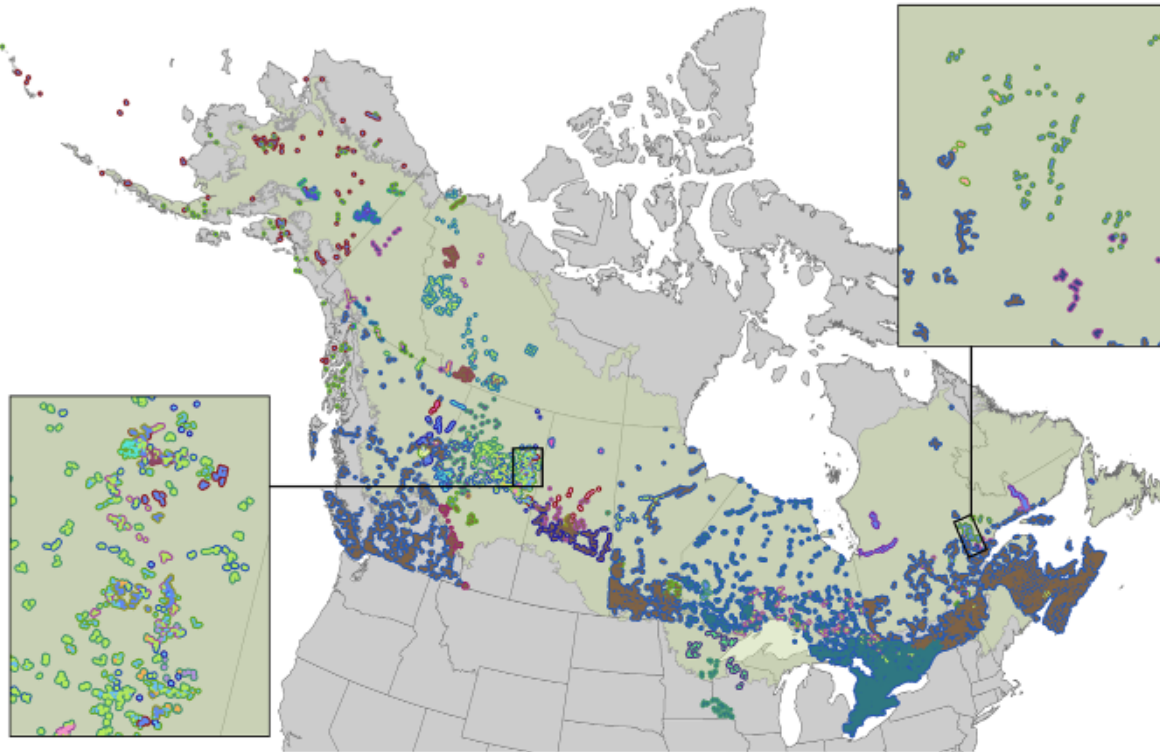
My protocol, your protocol

Figure 2: Barker et al. 2015

---

[^2]Ralph et al. 1993, Handbook of field methods for monitoring landbirds.

Survey methodology variation (colors) among contributed projects in the Boreal Avian Modelling (BAM) data base as of 2014[3].

Pop quiz

- In what regard can protocols differ?
- What drives protocol variation among projects?
- Why have we abandoned following protocols?

Moving away from standards

- Detection probabilities might vary even with fixed effort (we'll cover this later in depth)
- Programs might have their own goals and constraints (access, training, etc.)

Model based approaches

Less labor intensive methods for *unmarked populations* have come to the forefront:

- double observer (Nichols et al. 2000),
- distance sampling (Buckland et al. 2001),
- removal sampling (Farnsworth et al. 2002),
- multiple visit occupancy (MacKenzie et al. 2002),
- multiple visit abundance (Royle 2004).

Models come with assumptions

- Population is closed during multiple visits
- Observers are independent
- All individuals emit cues with identical rates
- Spatial distribution of individuals is random
- Etc.

Assumptions are everywhere

Although assumptions are everywhere, we are really good at ignoring them:

- Relativistic time dilation is negligible (as long as we are not on a space station)
- Samples are independent

---

[3]Barker et al. 2015, WSB 39(3):480–487.

Pop quiz

- Can you mention some other common assumptions?
- Can you explain why we neglect/violate assumptions?

The hard truth

Assumptions are violated in many ways, because we seek simplicity

The main question we have to ask: **does it matter in practice**?

Our approach

1. Introduce a concept
2. Understand how we can infer it from data
3. Recreate the situation *in silico*
4. See how the outcomes change as we make different assumptions

How to get away

It is guaranteed that we violate **every** assumption we make.

To get away with it, we need to understand **how much is too much**.

"All assumptions are violated, but some are more than others."–Me
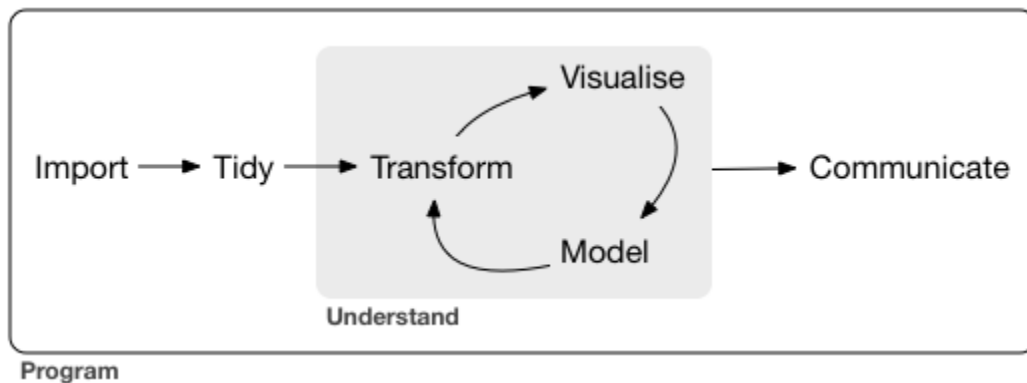
Tidy workflow

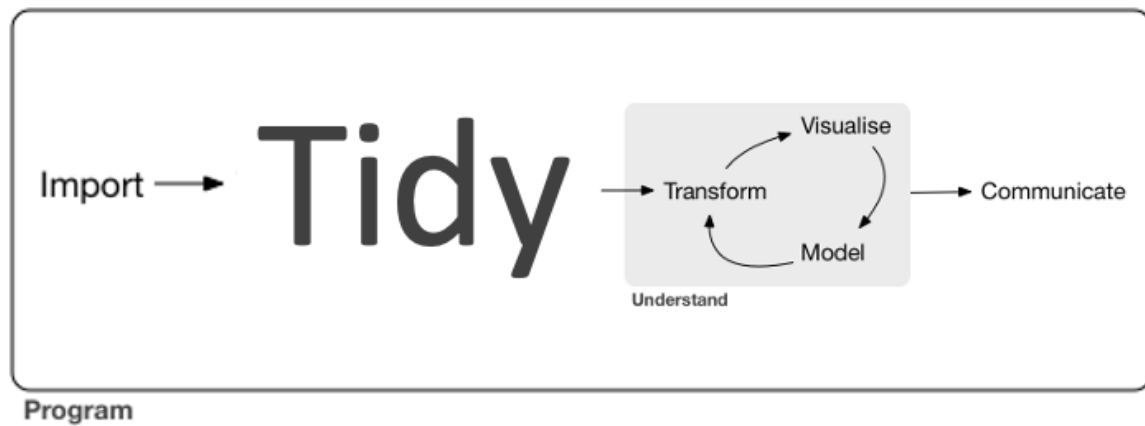Figure 3: Tidy workflow

Source:

Tidy time tracking

Figure 4: Tidy time tracking

Letters proportional to time spent

Data

It is often called:

- processing,
- compilation,
- munging,
- wrangling,
- cleaning.

None of these expressions capture the dread associated with the actual activity.

"All data are messy, but some are missing."

First mover disadvantage

If you are the first person to ever look at the data, hope for the best, expect the worst.

If no one looked, no one found the problems yet.

The four horsemen

Luckily, there are only 4 things that can get messed up:

1. space (e.g. wrong UTM zones),
2. time (ISO format please),
3. taxonomy (UNK, mis-ID),
4. measurement units.

Check out source code if you are interested in data processing, we skip that for now to concentrate on the fun part.
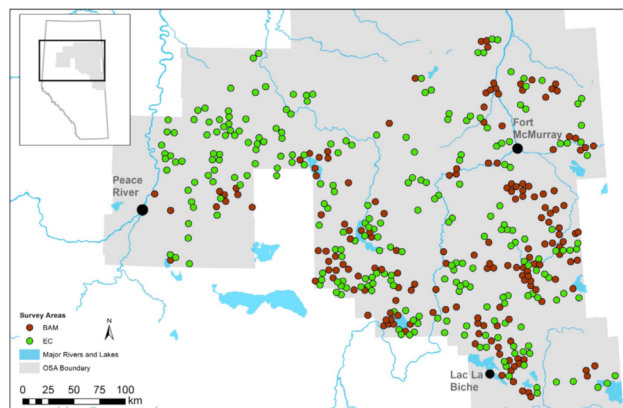
JOSM (Joint Oil Sands Monitoring) data

Figure 5: Mahon et al.

Cause-Effect Monitoring Migratory Landbirds at Regional Scales[45]: understand how boreal songbirds are affected by human activity in the oil sands area.

---

[4]Mahon et al. 2016, For. Ecol. Man.

[5]Mahon et al. 2019, Ecol. Appl.

Survey design
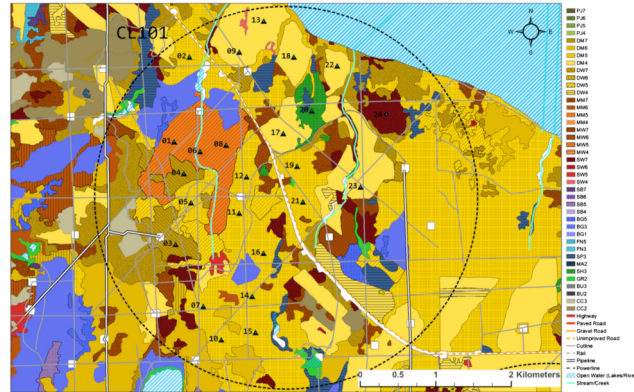

Figure 6: Mahon et al.

Survey area boundary (r=2.5 km circle), habitat type and human footprint mapping, and clustered point count site locations.

Sampling, replication

- We want to make inferences about a population,
- full census is out of reach,
- thus we take a sample of the population
- that is representative and random.
- Ideally, sample size should be as large as possible,
- it reduces variability and
- increases statistical power.

How do we pick survey locations

- Stratification,
- gradients,
- random location (control for unmeasured effects),
- take into account historical surveys (avoid, or revisit),
- access, cost (clusters).

Next

Overview of regression techniques