

Introduction

Peter Solymos

Point count data analysis workshop, AOS 2019, Anchorage AK,
25 June 2019

About the course

You'll learn

- how to analyze your point count data when it combines different methodologies/protocols/technologies,
- how to violate assumptions and get away with it.

About me

- Ecologist (molluscs, birds),
- pretty good at stats (modeling, detectability, data cloning, multivariate),
- R programmer (vegan, detect, ResourceSelection, pbapply),
- sometimes I teach (like today).

Install packages

```
pkgs <- c("bookdown", "detect", "devtools", "dismo",  
  "Distance", "forecast", "glmnet", "gbm", "intrval",  
  "knitr", "lme4", "maptools", "mefa4", "mgcv", "MuMIn",  
  "opticut", "partykit", "pscl", "raster", "sp",  
  "ResourceSelection", "shiny", "unmarked", "visreg")  
to_inst <- setdiff(pkgs, rownames(installed.packages()))  
if (length(to_inst))  
  install.packages(to_inst,  
    repos="https://cloud.r-project.org/")  
devtools::install_github("psolymos/bSims")  
devtools::install_github("psolymos/QPAD")  
devtools::install_github("borealbirds/paired")  
devtools::install_github("borealbirds/lhreg")
```

Download the book

The screenshot shows the GitHub interface for the repository `psolymos / qpad-book`. At the top, there's a navigation bar with links for Pull requests, Issues, Marketplace, and Explore. Below this, the repository name is displayed along with statistics: 1 Unwatch, 0 Stars, and 0 Forks. A secondary navigation bar includes links for Code, Issues (0), Pull requests (0), Projects (0), Wiki, Security, Insights, and Settings. The main content area features the repository description: "QPAD book: point count data analysis" with a link to <http://peter.solymos.org/qpad-book/> and an Edit button. Below the description, there's a section for repository statistics: 128 commits, 2 branches, 0 releases, 1 environment, 1 contributor, and a View license button. A progress bar is visible below these statistics. Further down, there's a section for recent commits, showing the latest commit by `psolymos` titled "Create qpad-book-slides-03-arus.Rmd" from 6 hours ago. Below the commit list, there's a table of files and their last update times.

psolymos / **qpad-book** Unwatch 1 ★ Star 0 Fork 0

Code Issues 0 Pull requests 0 Projects 0 Wiki Security Insights Settings

QPAD book: point count data analysis <http://peter.solymos.org/qpad-book/> Edit

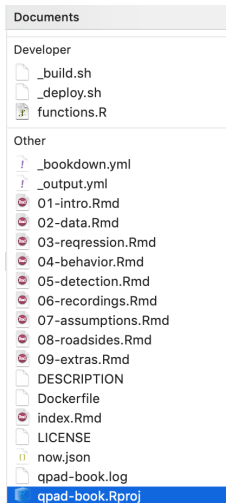
Manage topics

128 commits 2 branches 0 releases 1 environment 1 contributor View license

Branch: master New pull request Create new file Upload files Find File Clone or download

psolymos	Create qpad-book-slides-03-arus.Rmd	Latest commit e3a9de8 6 hours ago
_data	tweaks	6 hours ago
_scripts	edr issue solved	4 days ago
_shiny	Update bsims1.R	2 days ago

Open project



qpad-book.Rproj

R Project - 277 bytes

Tags [Add Tags...](#)

Created June 7, 2019 at 1:02 AM

Modified Today, 7:11 PM



More...

Notes

The screenshot shows a web browser with the address bar displaying `https://hackmd.io/@psolymos/aos-2019/`. The browser's address bar also shows a search icon, back, forward, and home buttons. The browser's toolbar includes a HackMD icon, a pencil icon, a document icon, an eye icon, a plus icon, a question mark icon, a 'New!' button, a share icon, a HackMD logo, and a '1 ONLINE' badge.

The main content area is split into two panels. The left panel is a dark-themed code editor showing the following text:

```
1 # Point count data analysis workshop
2
3 AOS 2019, Anchorage AK, 25 June 2019
4
5 Workshop material at
6 http://peter.solymos.org/qpad-book/
7
8 This document is hosted at
9 https://hackmd.io/@psolymos/aos-2019/
10 and contains notes taken during the
11 workshop.
12
13 ## 1. Introduction
14
15 ## 2. Organizing and processing point
16 count data
17
18 ## 3. A primer in regression
19 techniques
```

The right panel is a light-themed document viewer showing the rendered content of the document. It includes a 'PUBLISHED' status, a 'CHANGED 18 MINUTES AGO' timestamp, and a 'Watch' button. The document title is 'Point count data analysis workshop'. The content is identical to the code editor, but with the links rendered as blue text. The document is organized into sections: '1. Introduction', '2. Organizing and processing point count data', and '3. A primer in regression techniques'. The bottom of the right panel shows the start of a fourth section, '4. Behavioral complexities'.

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]

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

- You might say: I saw 5 robins (walking down the road),
- I might say: I only saw one (sitting on my porch)

Apples to apples

Effort:

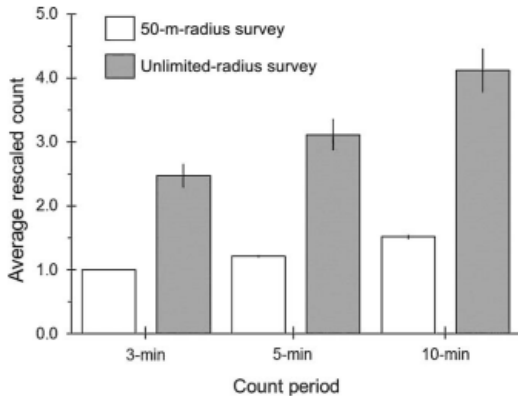
- area of the physical space searched,
- amount of time spent,
- number of individuals identified.

Experience, skill, “sensitivity”:

- number of years in field work,
- eye sight, hearing ability,
- mic sensitivity.

The goal is to make our measurements comparable.

Effects can be significant



10-min unlimited count ~300% increase over 3-min 50-m count.
Average across 54 species of boreal songbirds¹.

¹Matsuoka et al. 2014, [Condor 116:599–608](#).

So 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 abundance?

Standardization by design

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 counts².

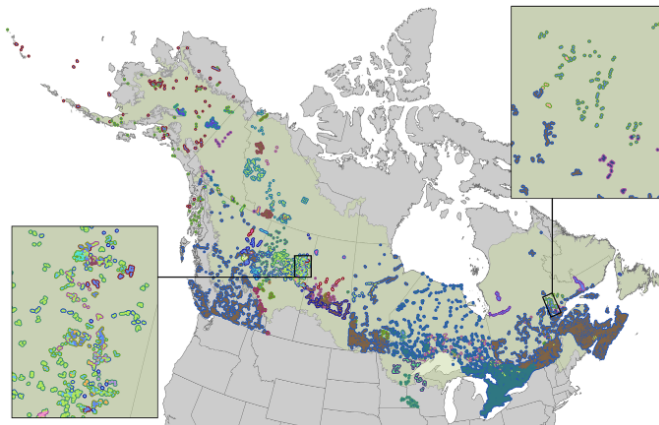
But programs started to deviate from standards:

“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.”³

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

³Matsuoka et al. 2014, [Condor 116:599–608](#).

Protocols do vary



Survey methodology variation (colors) among contributed projects in the Boreal Avian Modelling (BAM) data base as of 2014⁴.

⁴Barker et al. 2015, [WSB 39\(3\):480–487](#).

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 more later),
- programs might have their own goals and constraints (access, training, etc).

Model based approaches

Less labor intensive methods for unmarked populations has 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 uniform,
- etc. (we will investigate this further in depth).

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.

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. We will introduce a concept,
2. understand how we can infer it from data,
3. then we recreate the situation *in silico*,
4. and see how the outcome changes as we make different assumptions.

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.”

The rest of the day

1. Introduction
2. Organizing and processing point count data
3. A primer in regression techniques

Short break

4. Behavioral complexities
5. The detection process

Lunch break

6. Dealing with recordings
7. A closer look at assumptions

Short break

8. Understanding roadside surveys
9. Miscellaneous topics

Dismissal

<https://twitter.com/adamgruer/status/1122271095225118720>

Data ...

It is often called:

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

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

“All data are messy, but some are missing.”

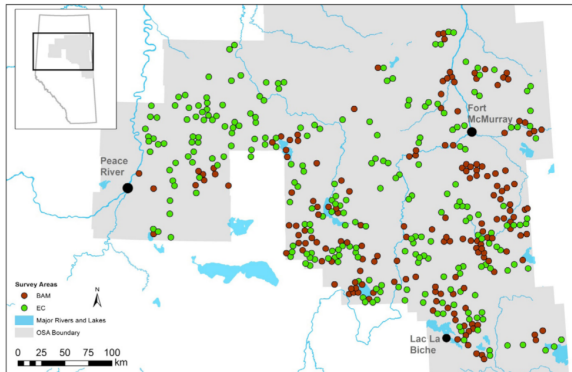
The four horsemen

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. something else (\rightarrow check again).

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

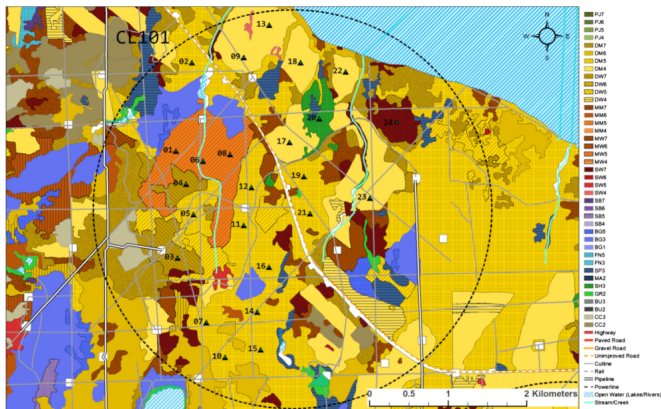


Cause-Effect Monitoring Migratory Landbirds at Regional Scales⁵⁶:
understand how boreal songbirds are affected by human activity in
the oil sands area.

⁵Mahon et al. 2016, [For. Ecol. Man..](#)

⁶Mahon et al. 2019, [Ecol. Appl..](#)

Survey design



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

Sample and 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 where to survey?

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

Dive into the JOMS data now

```
load("../_data/josm/josm.rda")  
names(josm)
```

```
## [1] "surveys" "species" "counts"
```