QPAD

Peter Solymos

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

Checking in

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

What is detectability?

In the most colloquial terms, δ is the probability that a species is detected given it is present:

$$P(Y > 0 \mid N > 0)$$

Occupancy

In an occupancy framework, we can have:

1. A detection ($Y > 0$), $P(Y > 0 \mid N > 0) = \delta\varphi$,
2. A non-detection ($Y = 0$), $P(Y = 0 \mid N > 0) = (1 - \delta)\varphi$,
3. An absence ($Y = 0$), $P(Y = 0 \mid N = 0) = 1 - \varphi$.

δ is the *false negative rate (FNR)*.

Abundance

A lot more combinations of true abundance and observed counts:

	$Y = 0$	1	2	...
$N = 0$	x			
1	x	x		
2	x	x	x	
...	x	x	x	x

Estimating detectability

To estimate δ, we need:

- ancillary information (multiple visits, distance bands, time intervals, observers),
- parametric model assumptions (i.e. δ varies across locations).

The myth of constant detectability

Detectability zealots often view a method that cannot estimate constant detection probability δ (e.g. single-visit occupancy and N-mixture models) as inferior.

Fortunately for the rest of us: δ can only be constant in very narrow situations, e.g. when surveys are conducted:

- in the same region,
- in similar habitat,
- in the same year,
- on the same day,
- at the same time,
- by the same observer,
- using the same protocol.

Constant detectability is rare

Often a consequence of small sample size (i.e. not a lot of detection for a species)[1,2]:

Removal models[a]	Distance models[b] (0) No effects	(1) TREE	(2) LCC	Total
(0) No effects	15	1	4	20
(1) JDAY	9	18	9	36
(2) TSSR	1	2	–	3
(3) JDAY + JDAY^2	1	2	2	5
(4) TSSR + TSSR^2	1	–	–	1
(5) JDAY + TSSR	3	4	–	7
(6) JDAY + JDAY^2 + TSSR	–	1	–	1
(7) JDAY + TSSR + TSSR^2	1	1	–	2
(8) JDAY + JDAY^2 + TSSR + TSSR^2	–	–	–	–
Total	31	29	15	75

Model ID and covariate effects [b]	M_0^φ	M_f^φ	M_f^c	Total	Residents
0: Null (time-invariant: M_0 and M_f)	1	0	0	1	0
1: DY	1	6	2	9	2
2: SR	1	5	4	10	2
3: DY + DY^2	2	3	4	9	0
4: SR + SR^2	0	7	17	24	4
5: DY + SR	0	5	3	8	1
6: DY + DY^2 + SR	2	0	1	3	0
7: DY + SR + SR^2	2	2	19	23	3
8: DY + DY^2 + SR + SR^2	1	1	4	6	1
9: LS	0	3	3	6	0
10: LS + LS^2	0	2	10	12	4
11: LS + SR	1	0	6	7	0
12: LS + LS^2 + SR	0	1	4	5	0
13: LS + SR + SR^2	1	3	18	22	6
14: LS + LS^2 + SR + SR^2	0	0	2	2	0
Total	12	38	97	147	23

[b] Covariate effects included linear or quadratic effects of time since sunrise (SR), ordinal day of the year (DY), and the ordinal day relative to local spring (LS).

[1] Solymos et al. 2013, Methods. Ecol. Evol. 4:1047–1058.
[2] Solymos et al. 2018, Condor 120:765–786.

Let's unwrap δ

1. Once the species/individual is present ($N > 0$),
2. it needs to make itself heard/visible, make itself available (p) – we have just discussed this,
3. then it needs to be detected by an observer (either in the field or in the lab listening to a recording) – let's call this probability q.

These imply a set total time duration and maximum counting radius.

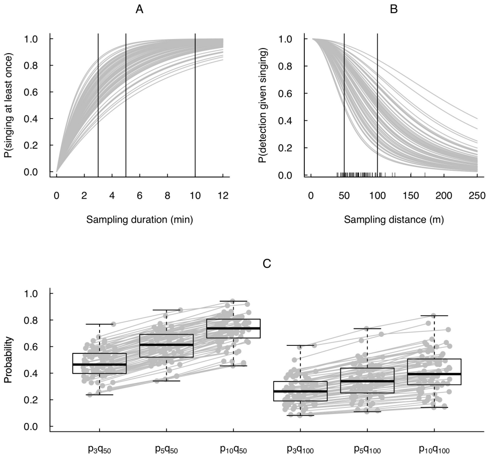
QPAD

Now we can expand our equation:
$E[Y] = NC = (AD)(pq) = qpAD$

The expected value of the observed count becomes a function of the:

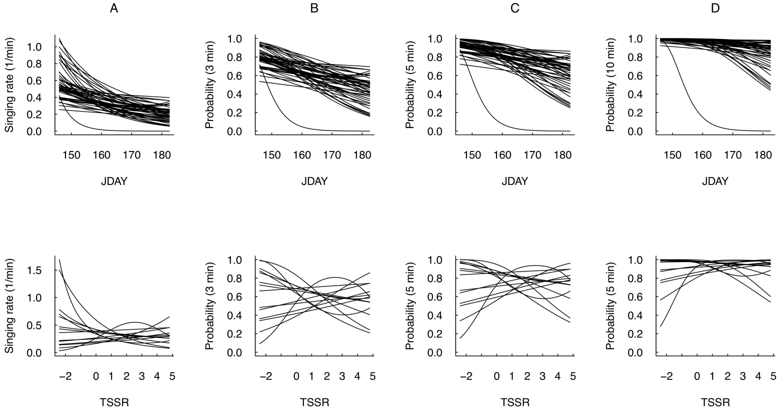
- population density (D),
- area sampled (A),
- availability (p),
- and perceptibility (q).

Space and time to the rescue

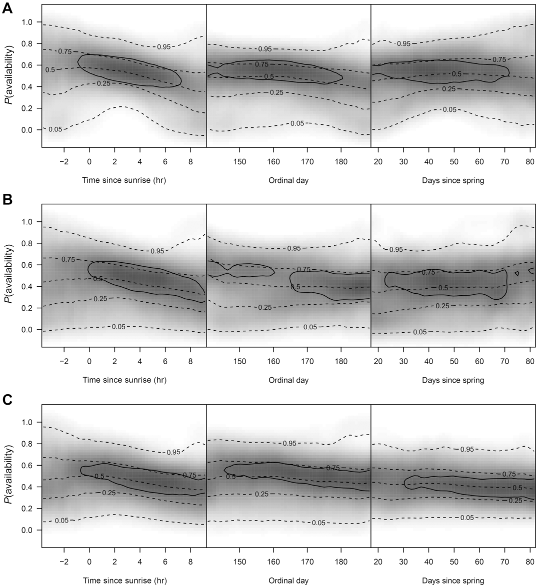


p is a monotonic function of *time*, while *q* is monotonic function of *area* (space).

Availability varies



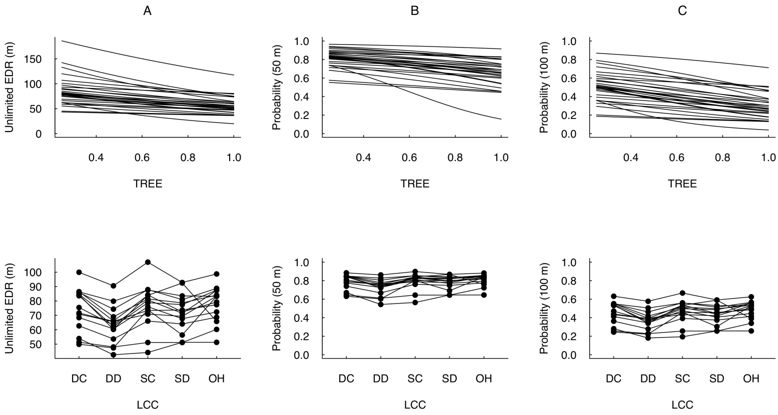
Availability really varies



Biological mechanisms

- Migration timing drives phenology for many species, e.g. ordinal day of year (`DAY`),
- when study spans across biomes, use time since local spring (multi-year average),
- or time since spring green up, last snow day, etc. based on actual survey year,
- time of day,
- time since local sunrise (`TSSR`).

Perceptibility varies too



Estimating nuisance variables

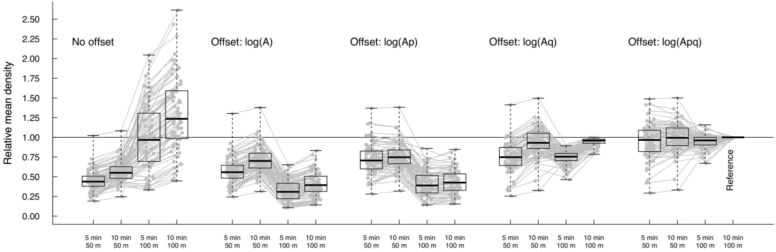
We have discussed how to estimate p.

Next we will discuss how to estimate q based on distance sampling.

So why is this important?

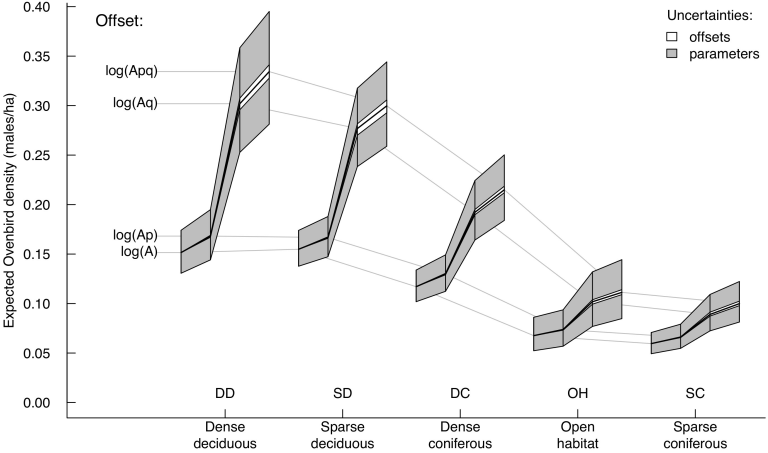
$$E[Y] = NC = (AD)(pq) = qpAD$$
$$\hat{D} = E[Y]/(A\hat{p}\hat{q})$$

Corrections



The corrections can be used to adjust for different methodologies and other factors influencing availability and detectability.

Ovenbird



Density estimates using different corrections in 5 types of land cover

Integrating projects

1. Normalize and join data from projects,
2. take multiple-duration subset and fit removal models,
3. take multiple-distance subset and fit distance sampling models,
4. predict p and g for all surveys,
5. use $log(Apq)$ offsets in log-linear models with total counts as response.

Removal and distance models can deal with multiple methodologies in the same model.

Each modeling step can include covariates, which can be the same.

Advantages of this approach

- Smaller parameter space,
- less colinearity,
- straightforward model selection,
- quicker than joint modeling via integrated likelihood.