Hi Peter,

Following up on our previous dialogue, I was hoping to get your thoughts on some results I’ve generated through simulations.

In all of these simulations, the observer only transcribes the first detection of each bird. Cue rate and EDR are identical for every bird and independent of each other, and I used a half-normal distance function to generate data and fit statistical models. All simulations assume 10-minute point counts broken down into 3/5/10 minute intervals, and using distance bins of 0-50m, 50-100m, and 100m+.

First I generated 1000 repeated simulations when tau = 1 (i.e., EDR = 100 m) and phi = 0.2 (i.e., cue rate = 0.2 per minute). Results are shown below. Interestingly, applying distance and removal sampling independently results in estimates of EDR that are positively biased, and simultaneously, estimates of phi that are negatively biased. The biases don’t cancel out when calculating density, and resultant estimates of density are positively biased.



Next, I repeated this process across multiple combinations of tau and phi. Below, I plot the median (from 1000 simulations) percent bias in density estimates for each combination.



In a third set of simulations (not shown), I found that QPAD estimates are unbiased if observers only transcribe the first cue produced by each bird. However, as we discussed before, it is hard to reconcile this with the logic of using detectability modeling in the first place: observers will fail to detect cues, and will never know if the first cue they detect was the first cue produced by the bird. All they can do is transcribe their first detection.

**Why are estimates of cue rate (phi) negative biased?**

The rate at which an observer detects cues from a bird decreases in proportion to the detectability of each cue, which is determined by the distance function; more distant birds will “appear” to have less frequent cues because many of their cues are undetected. Since removal modeling ignores distance, the “apparent” cue rates are being averaged across birds from a variety of distances. Distant birds, with lower “apparent” cue rates are dragging down the average.

**Why are estimates of EDR (tau) positively biased?**

This has to do with the “repeated cueing” component. Repeated cueing increases the “overall” detection probability of near birds only slightly (because they are already highly detectable from a single cue) but can dramatically increase the overall detection probability of distant birds. Since the distance modeling doesn’t consider cue rate and essentially assumes each bird only produces a single cue, EDR is getting positively biased because distant birds are entering the data more frequently than expected. This is best illustrated by an example:

Consider two birds at different distances from the observer: a “near” bird at 25 m and a “distant” bird at 125 m. Assume each bird produces 5 cues during a point count and EDR is 100 m.

The probability that a single cue is detected by the observer is:

P\_cue\_near = exp(-(25/100)^2) = 0.94

P\_cue\_distant = exp(-(125/100)^2) = 0.21

Each cue from a “distant” bird is about 4.5 times less likely to be detected by the observer. This what the distance modeling part of QPAD is trying to estimate. However, when each bird produces 5 cues during a point count, the probability that the observer detects each bird at least once is:

P\_bird\_near = 1-(1-0.94)^5 = 1

P\_bird\_distant = 1-(1-0.21)^5 = 0.69

Thus, overall, distant birds are actually only about 1.4 times less likely to be detected than near birds.

**Empirical confirmation**

If this hypothesis is correct, we should see the following in the empirical data (for data sets with time and distance bins):

* EDR estimated using all time bins will be higher than EDR estimated using only a single time bin
* Cue rate estimated using all distance bins will be lower than cue rate estimated using only the nearest distance bin

**Solutions?**

Model cue rate and EDR jointly: the likelihood for removal modeling needs to account for EDR (and distance of birds), and the likelihood for distance sampling needs to account for cue rate.

Unfortunately, I am not good enough at math/calculus/stats to figure out analytical solutions for the expected multinomial cell probabilities in each bin. I sort of (dimly) understand where we need to add parameters to the PDFs of for removal and distance sampling equations, but I don’t know how to integrate these across distances and times to calculate the CDFs (and ultimately the relative distributions of birds in each distance/time bin).

For example, I understand the PDF for the half-normal distance sampling function is:

exp(-(distance/tau)^2)

This