**FOR BOARD MEETING:**

**ANALYSES STILL TO DO**

* **Use the Lidar data (% forest cover in 500m). Perhaps smallest ring for detection, and largest ring for abundance??**
  + **BRMS\_GLMM**
  + **STAN\_DISTSAMP , WITH farthest 10% of detections removed (this may address the 100+m issue for many species?)**
  + **BBS?**
  + **DISTANCE, WITH farthest 10% of detections removed (this may address the 100+m issue for many species?). And then GLMM?**
  + **STAN\_PCOUNT— allows random effects**
* **Perhaps get the “better” sampling time covariate as time-since-sunrise**
* **NOTE: Center and standardize covariates**

**PRIORITY RESULTS TO SHOW - TRENDS OF COMMON SPECIES PER PARK**

* **For each park, plots showing the raw data and estimated trend with 95% CI on trend. Color red if statistically significant.**
* **For the subset of park-species for which detection can be estimated, show real data, estimated relative trend, and estimated absolute abundance/trend**
* **Across all parks and common species – a table showing species as rows and parks as columns. In the table, colored arrows showing trend and statistical significance.**
* **Show some kind of mapped results, perhaps per species show arrows of trends on the sites**
* **A table summarizing important covariates for each park-species’ trend and/or detection/availability**

**SECONDARY, IF HAVE TIME**

* **For species that violate detection/availability assumptions:**
  + **Which ones can would we need to combine first two bins for**
    - **What park-species seem to avoid observers? (or it could be due to different habitat in closest ring)**
  + **Which ones may have too large home ranges?**
    - **What species are likely detected way beyond the 100m?**
  + **And how much is this related to the understory density & heterogeneity rather than to species? Or coarsely classified habitat het?**
* **A summary of park-species for which each of the detection/availability assumptions more or less can be met.**
  + **How is this related to homogeneity of the site? Perhaps more complex, it may be related to how different % forest is in the different “rings”—bc if heterogeneous similarly across all distance classes it may be ok**
* **Perhaps show how site homogeneity or % forest is related to observed species trends and/or ability to meet detection assumptions? And show how site homogeneity differs across the parks.**

--------------------------

STILL TO DO:

* Finish Gmodels
* GLMM
* SUMMARIZE RESULTS with graphs incl detection function & availability function graphs against the actual data
  + GLMM
  + Stan\_pcount—rpt surveys
  + Distance--distance
  + Solymos – removal + distance >>>> CI’S NEED TO BE CORRECTED
  + Gdistsamp – distance + TE (rpt surveys)
  + Gmultmix – removal + distance + TE (rpt surveys) >>> DID NOT FIT WELL
  + Gdistremoval – removal + distance + TE (rpt surveys)
* Distance sampling estimates with just ONE survey per year—does it differ much from two surveys per year?

SUMMARIES TO SHOW FOR EACH SPECIES-PARK, WITH ALL ANALYSES

* Full AIC table with all analysis methods—SAVE AS EXCEL
  + Analysis approach
  + Species
  + Park unit
  + Data structure (.Rdata file)
  + Notes of issues
  + Included covariates (i.e., model formula)
  + Mixture (P, NB, ZIP)
  + Keyfun (for distance models)
  + Mean overall bird density (across all sites and surveys)
  + Trend—is it significantly positive or negative? (could give actual value & CI but it will be scaled and need to figure out how to get backtransformed CI’s)
  + Model fit
* Plot distance detection function against actual data (for distance models)
* Plot availability against actual data (for removal models)
* Plot estimated density & 95%CI per plot per year (against real data)

SUGGESTIONS:

* Continue collecting distance and time-to-removal data
* But possibly cut off sampling at 100m (unless want to use the other data for other reasons, or if it may then bias the 100m data). For some species may want to keep the infinite distance data (e.g., AMCR) and simply estimate abundance indices
* Using unlimited radius—could be problematic when NOT doing distance sampling b/c, e.g., would underestimate in forests and overestimate in open spaces.
* Be more careful about sampling for those species where we are seeing a much lower detection probability (based on raw data) at the closest distance bin—not sure what to do about those because if we combine the first two distance classes, we will only have two known distance bins and we can’t fit a curve to that (would have to look into other options. Solymos EDR, but be careful of methods that promise something for nothing)
* Focus on getting better covariates to improve ability to estimate. Really, we are only going to see a big benefit with N-mixture models if we have good covariates for detection that actually parse out detection noise.
  + Decibel meter for noise? (but then how will we use the old category data—so maybe they should record the category first, then do decibel meter. This will help also for when observers change.)
  + Have the experts discuss other covariates (incl. ones not yet collected) for abundance (among sites) and detection.
  + Habitat covariates like % cover forest within 200m, 400m… so perhaps we can model across parks but with park and observer as fixed effects

BIGGEST DIFFICULTIES WITH THE DATA:

* Time-to-detection models are not fitting. May need to wait until we have more data so we can fit the 2-rate models. Perhaps once we get a standard habitat covariate we can combine data across Parks and just include Park as a fixed effect covariate for abundance. If we do this, will need to really standardize covariate estimates (e.g., noise) across the observers or it will be confounded.
* Distance models seem to be fitting w/100m. cutoff but not in all cases. Problem areas:
  + Birds that avoid the researcher, so would need to combine first two distance classes.
  + Sites with nonrandom heterogeneity within or immediately around the point count area can bias the distribution of birds w/respect to distance from observer.
* Assumptions of no double counting and of closed populations

WHAT CAN/SHOULD WE CHANGE IN THE FIELD?

* If we can fit the distance or removal part of the data adequately, would not need the repeat samples so much…
* How much time are they sitting after arrival to allow things to “calm down” (might help with both the distance and noise model fitting)
* Can we get a 100-150m distance class, then 150m+? If that is possible, we can use that information to see if for certain species we can “safely” say that in historical data most of the detections in 100+ is actually in 100-150m distance class. This will then make it possible to fit distance detection function to the people-avoiding birds.
  + Otherwise we can’t estimate distance b/c don’t have a known denominator
  + At the moment wouldn’t put too much stock in EDR b/c our data are binned distances so any estimate of EDR on just a couple distance intervals will be poor
* Mark if the detections are visual rather than auditory. Really those numbers are going to be quite low so it would be best if for visual detections they can also record distance and time of first auditory so we can actually use that information (but if we use the data for just pcount can use the visual even if never an auditory detection in the 10min, and it would actually be useful to know what % of birds are detected visually but never auditory and how much they moved btwn the two detections. Perhaps mark as V1 then if there is a later auditory, mark that as V1)

OTHER THINGS TO LOOK INTO WITH ANALYSES

* What other analyses can work without repeat samples?
* What can we do to make it reasonable to analyze data across Parks?

BIGGER PICTURE THINKING

What are we really worried about with the birds? We don’t want them to decline without us detecting.

Do we really need repeat visits then? Well we want to make sure a trend in detection does not obscure our ability to detect a decline

What do we need for that?

* We need to make sure we have good covariates for detection probability. Better habitat and other covariates

Repeat visits is time consuming, what is the trade-off? Can we get better bang for the buck doing something different?

* What about increasing the number of sites in each park and going to each site every other year, but repeat visits. But 1/3 of panel is visited every year.

Taking advantage of BBS data from around the park and putting park trends in context. bbsBayes etc.—we can use the NPS data the same way, do they think any of their observations were beyond 400m?

We should have a frank discussion with the other networks about their analysis experiences and whether Nmixture models with repeat samples really are the best option or if we should consider other methods (e.g., USGS just does Bayesian glmm’s). It seems that the folks using Nmixture models aren’t necessarily using it because they think it’s amazing. Worth talking to the folks who have done custom Bayesian modeling, but also talk about just moving toward the USGS approach?

How much do the detection and availability estimates actually change density estimates? Matt said it’s all just parallel but higher and with large CI’s. JP said their numbers are sometimes all over the place, and that may be b/c of model misspecification. Jeff Doser (NETN paper, but Kate Miller disapproves) has done analyses. My thinking is that with 10-min point counts, availability is probably already high (most birds there will have called) but detection probability may be more the factor that will potentially have covariates (e.g., habitat) that confound with abundance. So if anything—even though more difficult to collect well—perhaps keeping the distance-to-detection information is most important (but can we get a 100-150m distance class, then 150m+?)

Typical covariates for availability are julian date and start time. The effect of julian date is likely to have a long-term trend itself with climate change. Really we are not at much risk of having a trend in availability due to these cov (at least not in a way we can “fix” with analysis), which is another reason the time-to-removal data may not contribute much beyond a multiplier on the raw data.