**MY ANALYSIS THOUGHTS, FROM READINGS**

**Analyses to try:**

**CENTER AND SCALE COVARIATES**

* Basic GLM (Bayesian)
* Solymos QPAD offset GLMER approach, with bootstrapping—try with and without unlimited distance. Ideally would use 2-level mixture model for “high” and “low” frequency singers but not enough sample size
* Chris Ray approach??
* JP approach

**Models to try:**

* **For estimating from removal model**
  + Null or 2-point mixture with these model combinations >>
  + No cov and all combinations of these >>
    - Time-since-sunrise (quad)
    - Ordinal day
    - General habitat category, for parks where seems relevant and based on EDA
    - Noise or wind or other covariate—based on EDA

**Propagating uncertainty:**

* [From Solymos suppl]To incorporate uncertainty w.r.t. singing rate and distance sampling parameter estimates, one can use nonparametric bootstrap to generate o\_sets. These o\_sets will represent the uncertainty in singing rate and distance sampling parameter estimates, so the error can be propagated through the GLM.
* [From Solymos suppl] It is also possible to take into account model selection uncertainty in the QPAD estimates by randomly choosing among the possible singing rate and distance models based on model weights (with and without the parametric bootstrap procedure in the o\_sets switched by the boot argument).

**Covariates to still work on**:

* Some estimate of % tree cover within 100m or EDR
* Time-of-day RELATIVE TO SUNRISE, as a quadratic effect (SR + SR^2) – combined with ordinal day of year? (R package maptools to get local sunrise and spring)
* How much does Julian date change? B/c Solymos found it to be an important covariate but maybe bc working with data spanning large range of Julian dates
* Some measure of edge habitat, or something like yes-no road within the 100m point count radius?
* Possibly have 2-point mixture removal model where % infrequent singers stays the same and perhaps the singing rate of each group stays the same, or just the singing rate of each group changes with habitat etc covaraites but the % infrequent singers stays same? (maybe will have enough data for at least this 2-point model)

**Other notes:**

* Solymos 2018—species needs to be detected in at least 45 point counts for 90% probability of successfully converging on project-level estimates. This was estimated per project-species. For Mf model (2-point?), need detection during 158 point counts [and to successfully estimate the associated variance of the model parameters, need N > 963!!]
* Solymos 2018—they fit the conventional and finite mixture models to species with at least 75 non-zero point counts
* I should compare parameter estimates when using data since 2019 vs all data to estimate the detection parameters
* Solymos 2018—interestingly, for species where Mf was better fit, 75% were better when % infrequent was allowed to vary over time vs. when singing rate was allowed to vary over time
* Solymos 2018 (and consistent with other studies)—high detectability 0-4hrs after sunrise and during ordinal days 155-175
* Solymos 2018—in general, it seems we need at least 200 non-zero point counts for the species to do model selection
* MUST confirm that nearly all detections are auditory—otherwise, it messes up the estimation. In fact, in their datasheets it may be better if they only record first auditory detection (though if visual they can record it as visual—then when they first hear that bird sing, enter it again (on same record). OR just record if it’s visual and those can be analyzed separately but highly unlikely they would have enough sample size to work with it

Add to dashboard:

* Total number of non-zero point counts (I think I have this already)
* Histograms of # of surveys by time-of-day and day-of-year, with lines for the “core” (best) periods of 0-4hrs after sunrise and ordinal days 155-175
* Figures looking at non-zero counts as a function of time-since-sunrise and time-since-spring-start (and the related ordinal day—actually for now the time-since-spring may not matter bc they were using 30-yr averages); also look at median count as the metric

**6/27/22 Chats with Whitney about longer span of data**

we have 10 more years of data for JELA, VICK, and SAAN but it does not include the double sample design... do you think that data could be useful?

Yes! The only difference is it wasn't double sampled, but otherwise all done the same? That would definitely help (the 10 more years of data)--if nothing else, it could help us get better estimates for detection and the various covariates.

yep, same just no double sample.

[1:34 PM] Granger, Whitney

ok, I forgot that the parks where on a rotating panel design back then set up by Dan Twedt. that means not every point was sampled every year. are you still interested in that data?

[5:04 PM] Granger, Whitney

each park will have different date ranges of data because of the way the projects came on line

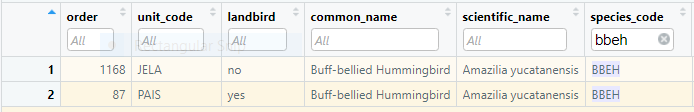
[5:06 PM] Granger, Whitney

also, after the 2019 protocol re-write, some of the points were reduced to a smaller number so there will be some data pre 2019 for some locations that we currently no longer sample.

**6/13/22**

[Response: BBEH should be changed to YES landbird for JELA. PAIS is not being sampled b/c they only ever detected 3 species of birds.]

Teams message to Whitney checking on this, seems incorrect



**6/9/2022 phone call with Whitney**

Whitney said the “breeding” designation is by species and does not indicate if a particular species is breeding in a particular park. BUT when I look in the species data frame, it seems that for any species it may be listed as yes or no breeding depending on the park.

**6/6/2022 email to Whitney**

[Response: Whitney checked paper data and that really is the date they entered on the paper data]

Also, would you mind checking on this--

GUISMS32 is the only point in GUIS-MS that was surveyed on 5/28/2021 instead of on 5/8/2021. Is there a chance the date was entered incorrectly in the site conditions table and it should actually be 5/8/2021?

**Response to 6/6/2022 email**

* The observation records with count = 0 should be 1 –now fixed
* Full site conditions file provided (some didn’t export last time)
* Issue 5.  These records are legitimate blanks for the time\_bin.  No time was recorded because they where detected before or after the count began and where not detected during the actual count, according the the detection notes. I guess Dan figured they were still worth mentioning in the database based on the species.  For this analysis, those records should not be used since they were not recorded during the allotted sample time frame.

**6/6/2022 email**

Hi! I'm not sure who I should be cc-ing on bird data questions so just let me know if I should be adding Jane and Whitney to these emails.

When you have a chance, could you look at items #3, #4 and #5 in the attached html of possible bird data issues and make any changes if necessary (for example, item #4 indicates that two observation records have count = 0, so please check if that is correct (b/c typically 0 counts are not entered, right?) or if it was a possible mistype.

Item #1 is not important, but if you want to update and send the revised parks list you can. It's just saying that these birds were reported in the bird observation file, but in the parks list they are not reported as having been reported in a bird count--so that should just be updated to 'yes' reported in a bird count (if I'm understanding it correctly). Again, not important for my purposes, just a heads up for you.

Thanks!

Ellen

**6/1/2022 email**

Hi, for the purpose of park summaries, do you want GUIS-FL to be distinguished from GUIS-MS?

In the park species lists (the ones that say what birds are found in what parks), they are listed as separate parks (in the parks column). In the bird observations, they are combined in one "unit\_code" (which, except for GUIS, seems to be equivalent to the "parks" column in the park species lists).

**5/31/22 email & response**

Ellen -

Here is some more information about vegetation data that Jane pulled together regarding locations that overlap our bird point areas. Since we are in the first stage of establishing vegetation monitoring plots we do not yet have data for all of the parks. Here is a summary of where they do overlap and we have data. That being said, we would have to have a conversation about which vegetation data would be most applicable to use:

There are about 23 points that overlap, with the best representation at BITH (Turkey Creek only, 10 pts), SAAN (4 pts at Rancho only) and PAAL (9 pts).

Here is an easy link to the vegetation community maps that cover all of the parks. Links to all the completed projects are on our website at the bottom of this page <https://www.nps.gov/im/guln/inventory-reports.htm>

If you need help compiling these resources, Jane should be available but not for a few weeks. She has a lot on her plate right now.

Hope this helps!

M

Great that would be wonderful thanks! Even if the data aren't available for all the points, I think it would be useful to see if there is any obvious relationship btwn the habitat/veg data and bird detectability/abundance (by species). I would expect the answer would be yes, for some species with enough data, if there is quite a bit of variability in habitat type among points for some parks. And in that case, including such habitat data (it might mean trying to find some quick/coarse way to characterize it for all points) could really improve the precision of detection--and therefore, abundance--estimates.

Ellen -

We happened to be having a staff meeting when you email came in so we had a chance to talk about this. Short answer is that you are correct, the bird locations were not filtered by habitat type. We do not currently include a habitat description as part of the data collected in the bird protocol. And yes, there are cases where the veg plots are near bird points but the number of sampling locations are not the equal so we would not have vegetation data for every bird point.

All that being said, Jane is going to look into what kind of vegetation community data we have that could be linked to the bird points. This could include the vegetation community maps and potentially some lidar data.

More to come...

Martha

Hi Martha,

For the birds point counts data, do you also record a general "habitat type" for each point or anything like canopy density/cover? I'm just trying to think of potential additional covariates I can use for detection times/distances. From the protocol I recall that the points can fall in various different habitat types within a park (but let me know if I'm wrong on that)? I thought that the bird data were often paired with vegetation data collected at nearby points or something like that?

Thanks,

Ellen

**3/23/22 phone call with Martha**

Comments:

* Bird density in park. Don’t know how much of the park is considered sampleable habitat. For example, dense brush, steep terrain, etc. are all not sampleable and may have bird densities different from sampleable areas. So estimates of density will need the caveat that it pertains to sampleable areas. Some sites are sampleable but not accessible. Useful to know what % of randomly located sites dropped for various reasons.
* Initial random draw is for all terrestrial habitats, whether it’s forest or non-forest, right?
* Exclude areas within 50m of road, building, etc. – so for edge species, potentially biasing park estimates low and would need to make that explicit; making inference to areas that are beyond that buffer
* Double observer samples? Use to assist with estimates of detection probability

Meeting:

* Time and distance
* Now repeat sampling but still collecting time and distance (doubled cost)
* Can we look at the data and see if we would actually come up with a different status report using repeat sampling.
* If we have to do repeat sampling, we want to know if we need to do time and distance.
* QAQC data is with experts.
* Occupancy, relative abundance, and density (last is only for most common)

**3/23/22 email from Jane**

(I had requested the breeding landbirds lists)—THEY MAY NEED TO BE UPDATED PERIODICALLY

Hi Ellen,

I created such a list for every park back when we were finalizing the protocol.  I've attached them all here, as a word doc for each as well as a single excel file for all parks.

These docs haven't been updated since the protocol was written, so the one you have for Palo Alto Battlefield is the same as what I've attached for that park.  There may be a few new species since 2016 that would need to be added to update these docs, but most likely just to the GULN detection list not the underlying park-wide list taken from NPspecies.

The exports from the bird database also include a column called "Breeder" which is based on that same lists.

**3/17/22 email from Jane**

Hi Ellen,

I heard yesterday on a group call that SIEN had just completed their first round of trend analysis for birds, and it looks like they use an occupancy type approach. Here’s a link to their protocol, tho I bet it doesn’t say much about what this as-yet-unpublished trend report does. <https://irma.nps.gov/DataStore/Reference/Profile/2124954>

<https://www.nps.gov/articles/sien-birds-overview-brief.htm>

I don’t have a direct contact with whoever would be able to familiarize you with what they did, but I do have a standing invite to set up a meeting with Mark Johnson at MIDN to get the bird analysis perspective from the northeast. I’m sure the same invitation is open to you without me attending, if you just reached out and set something up with him.

As Martha will probably discuss with you, even though the first set of reports for us will not analyze trends (just estimate population density for the most common species), Martha is really interested in getting some clarity on whether she can eventually phase out collecting data for both distance-and-time as well as repeat counts for occupancy. So she’ll probably want to see headway on the n mixture models in this first round, in addition to some version of the distance-and-time offsets, like those described in our birdSOP06.

Take care,

Jane

**10/20/21 phone call with Jane**

Need status report but not trend report yet

Priority objectives:

* Objective 2: Determine occupancy and the relative abundance of all species (per monitoring point) and the density of each park’s commonly detected species (per hectare). A species is considered commonly detected if it is recorded more than 40 times per year for at least three years. [EC: density per hectare—depends on how they sampled. I assume they didn’t just put random points in park, but stratified by habitat? commonly detected – shouldn’t it depend on the number of sampling points in the park in the habitat type that is of importance to that species?]
* Objective 3: Determine long-term trends in occupancy and relative abundance or density of commonly-detected species at each park.

Develop methods starting 2019, then lower priority decide how to use the older data.

Jane will send SAAN data –start with that one. Also some old R script and she will see what Great Lakes or other groups are doing for their bird analyses.

Ellen will be wrapping up cobble bar and bat stuff, then birds will be prioritized next—probably starting around end of this year (2021).

**8/13/21 email from Jane with previous correspondences with Dami**

**From:** Carlson, Jane E <Jane\_Carlson@nps.gov>  
**Sent:** Wednesday, December 16, 2020 3:01 PM  
**To:** deyitayo <deyitayo@uga.edu>  
**Cc:** Granger, Whitney <Whitney\_Granger@nps.gov>; Segura, Martha <Martha\_Segura@nps.gov>  
**Subject:** Re: [EXTERNAL] Re: GULN Bird Project

Hi Dami,

It was very nice meeting you on the Teams call today.  Our group is excited for your participation, and I look forward to helping you get familiar with our bird monitoring data. Since this isn't one of the protocols I'm directly in charge of, I had to spend quite a bit of time several years ago, unravelling the condition of the datasets. In sharing with you these notes and additional documents, I hope to spare you some of the same time and effort (although you'll still have to wade through this lengthy email-- my apologies in advance).

The first bird dataset I'm sending is for Palo Alto Battlefield National Historical Park, which is in Brownsville, Texas. Of the 6 parks where we collect bird data, it's the only one that had all points sampled every year, rather than having an alternate-year panel design prior to 2018. Palo Alto had the same 29 points sampled each year, starting in 2011 and ending in 2019. For 2019 alone, each point was sampled on two different occasions within the same season, using the same 10-minute point count methods. The dataset for this park should continue in summer 2021, but in 2020, we were unable to get the cooperator out in the field due to COVID travel restrictions.

Palo Alto is also a useful starting place to explore analytical approaches because it's the same park I used as an example in the data analysis SOP you've already looked at (linked directly [here](https://irma.nps.gov/DataStore/DownloadFile/607543)). I still have on hand the R scripts and underlying .csv files that I used to calculate detection probability estimates from the distance and time data, following Solymos 2013 paper cited in the protocol SOP (also attached, plus a few others) . The R scripts are not annotated or updated since 2017, though, and if you will not be using the distance and time data at all, they would not be of much use.

Our network team retained distance and time approaches in the protocol because the first avian ecologist involved selected those methods, and there still exists professional support for it, based on my own review of the literature that you'll see in the SOP (admittedly, not deep).  Based on Brian Mitchell's comments, you may be focusing primarily on the occupancy modelling and N-mixture models, but our staff remains interested in some independent consideration of the validity of a distance and time approach versus the occupancy approach. If there was indeed strong quantitative evidence that that distance and time data provided unreliable results or could only be used on a very few species, then we may eventually want to remove it from the protocol altogether.  But in the interest of time, we consider it the greatest priority to address our protocol's second and third objectives using (any?) analysis approaches that are consistently reliable and scientifically defensible.

A few more details about the dataset: the data export includes all bird species, passerine or otherwise, recorded during the point count event.  Our protocol focuses on breeding landbirds only, and the attached word document, called Palo Alto Species List, provides the appropriate classification for each species on the park's species list.  Similar information is in Table S6.1 in the data analysis SOP. This means that even though we had over 100 detections of northern Bobwhite in 2011-2016, we would not include it in our breeding landbird reporting.

You'll noticed there are three sheets within the excel document, which includes the database export.  One sheet is called Observations, for the individual observations of bird species during each point count (1 row for each bird species within a point count event), the second is called Site Conditions (1 row for each point count event), and the third is called Plot Locations (1 row for each site, to provide the GPS location).  The Conducting Bird Surveys SOP (linked directly [here](https://irma.nps.gov/DataStore/DownloadFile/607539)) would explain what the codes in Site Conditions mean.  The codes are recorded primarily as ancillary data, which could be used to account for environmentally-caused differences in bird detectability across time, though we haven't spent much time considering them or testing their utility.

As you'll quickly see, the observation dataset is not formatted correctly for occupancy type analyses, in part because it includes no non-detects, or zeros. The assumption is that if the observer did not record a species as present, it is a non-detect, so you can generate the zeros yourself.  Also, if there was a point count event where the observer did not see or hear any birds, that point count event would not be represented at all on the observation sheet, though it would be in the site conditions sheet. I'm not sure if that happened or not for Palo Alto, but it was something I ran across in certain parks.

Another attachment I've included is the supplemental materials document to the protocol, which describes the transition to from pilot study to final design.  If you notice inconsistencies over time within a given park's dataset, this would be the place to check first for an explanation (or just email me or Whitney, the network's data manager who produced the export I've attached; he's cced on this message and will participate in future meetings).

A shortcoming of starting with the Palo Alto dataset is that there is only one year of repeat-sample data. We are still in the process of quality-checking the 2020 data for the remaining parks.  After QA/QC is all done, I'll send you data for a second park, which would have two years of repeat-sample data, plus some pilot data.

Finally, I mentioned during our meeting the occupancy analysis approach I just learned about last week, where occupancy is modelled based on just one visit per season, rather than repeat-visits, using time to detection. The person who presented this was Brian Halstead, and I've attached his 2018 paper, in case that might be of use.

Kind regards

Jane

**From:** Damilola Eyitayo <deyitayo@uga.edu>  
**Sent:** Thursday, December 10, 2020 2:40 PM  
**To:** Segura, Martha <Martha\_Segura@nps.gov>; Mitchell, Brian R <Brian\_Mitchell@nps.gov>  
**Cc:** Carlson, Jane E <Jane\_Carlson@nps.gov>; Cheng, Ellen <ellen\_cheng@nps.gov>; Brian Irwin <irwin@uga.edu>  
**Subject:** [EXTERNAL] Re: GULN Bird Project

|  |
| --- |
| **This email has been received from outside of DOI - Use caution before clicking on links, opening attachments, or responding.** |

Thanks Martha. I'll reach out if I encounter any problems.

Best regards,

Dami

Damilola Eyitayo, Ph.D.

Postdoctoral Research Associate

Georgia Cooperative Fish and Wildlife Research Unit

Warnell School of Forestry and Natural Resources

University of Georgia, Athens GA

Phone: 740-856-5136

**From:** Segura, Martha <Martha\_Segura@nps.gov>  
**Sent:** Thursday, December 10, 2020 3:38 PM  
**To:** Mitchell, Brian R <Brian\_Mitchell@nps.gov>; Damilola Eyitayo <deyitayo@uga.edu>  
**Cc:** Carlson, Jane E <Jane\_Carlson@nps.gov>; Cheng, Ellen <ellen\_cheng@nps.gov>; Brian Irwin <irwin@uga.edu>  
**Subject:** Re: GULN Bird Project

[EXTERNAL SENDER - PROCEED CAUTIOUSLY]

Dami -

Here is the link to the landbird monitoring protocol narrative: [DataStore - Published Report - (Code: 2256606) (nps.gov)](https://irma.nps.gov/DataStore/Reference/Profile/2256606" \t "_blank)

And the link to the SOPs, including the analysis and reporting SOP: [DataStore - Standard Operating Procedure - (Code: 2255922) (nps.gov)](https://irma.nps.gov/DataStore/Reference/Profile/2255922" \t "_blank)

Please let me know if you have any problems downloading the documents. As a heads up, I believe the DataStore will be off line for a few hours this afternoon to install updates. You may want to wait until tomorrow to access the site.

We look forward to working with you on this!

Martha

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Martha Segura

Program Manager (Inventory and Monitoring)

National Park Service, Gulf Coast Network

646 Cajundome Blvd., Room 175

Lafayette, LA 70506

Tel: 337-291-2113

Cell: 337-257-0432

Fax: 337-291-3078

[martha\_segura@nps.gov](mailto:martha_segura@nps.gov)

<https://www.nps.gov/im/guln/index.htm>

**From:** Mitchell, Brian R <Brian\_Mitchell@nps.gov>  
**Sent:** Thursday, December 10, 2020 1:25 PM  
**To:** deyitayo <deyitayo@uga.edu>  
**Cc:** Segura, Martha <Martha\_Segura@nps.gov>; Carlson, Jane E <Jane\_Carlson@nps.gov>; Cheng, Ellen <ellen\_cheng@nps.gov>; Brian Irwin <irwin@uga.edu>  
**Subject:** GULN Bird Project

Hello again Dami,

The Gulf Coast Network bird monitoring project is the other project I’d like to get on your list. My understanding is that Jane would like some assistance with a trend report that is scheduled for this fiscal year. I believe these data are collected at point count locations using the removal method (equal time intervals where the time when a bird is first detected is noted). These data should also be suitable for occupancy modeling.

Can you please set up an initial meeting with Jane, Martha, Ellen, and me (and Brian I. if he wants to participate)? I think we should talk about what the network needs and what timeline is feasible.

Jane or Martha, can you please send Dami a copy of the protocol?

Thanks,

Brian

20201216 – GULN Birds project

Jane Carlson

Ellen Cheng

Dami Eyitayo

Brian Irwin

Brian Mitchell

Martha Segura

Brian M.: I spoke to Martha earlier today, and the pressure for this project is not as high as initially thought. GULN is aiming for a trend report in FY2022. So this project is probably third priority on Dami’s list, after CUPN bats and APHN cobble bars.

Dami confirmed that he has the protocol and SOPs.

Jane: I wrote the analysis and reporting SOP and am familiar with the data, but am not the protocol lead. Data collection started in 2011. It was a panel design. 1 park had full (all points, one time) sampling though 2018, and 5 had half points each year (2 panels). In 2018 GULN sampled all points at all 6 parks (no panel then or moving forward). In 2019, GULN began to visit each point twice. The data methods have not changed over time, nor have the points. May make sense to start with PAAL where all points were sampled every year. There was a document posted with the protocol that covers the changes in the design over time.

Dami: How do you want this analyzed? Park-specific or pooled? Jane: Parks will want individual analyses, but there is potential for interesting multi-park analyses. Martha: The parks cover a variety of habitats and there is perhaps not much we can do network-wide.

Dami: Looking at species numbers, relative abundance. The protocol suggests N-mixture models, but says they are still being developed. Change in sampling design allows use of occupancy modeling.

Brian: My post-doc was on occupancy and removal models, so I have experience with them. N-mixture came along later, and it’s been a while since I reviewed those papers. I think N-mixture is being used by CUPN.

Ellen: Unmarked in R can implement N-mixture models, and was developed by someone at UGA who is pretty responsive. I used it to analyze the GULN herp data.

Dami: When can I get a look at the data?

Jane: I can get data to you any time, though we’re still working on data from the 4 parks we sampled this year. So there’s only one year with full data (all points and repeat sampling).

Martha: Maybe start with PAAL and SAAN; they are the most data-dense of our parks.

Brian I.: I’m familiar with removal for fish populations, not sure if this is the same. Can you explain it?

Brian M.: Removal modeling at bird points is done mentally, not physically. Birders record the first time they hear an individual, and then subsequently ignore that individual. The difficulty is that the birds may move around the site, and it is a judgement call to decide if a song from a new location is really a new individual.

Jane: At the science review I attended for Greater Yellowstone Network last week, they talked about an occupancy model that works for one occasion.

Brian M.: I worked with similar data for my post-doc, where we collected repeat counts with a short break in between. The problem is that the counts are not really separated enough for independence, which can make it harder to interpret the results.

Dami: What’s the suggested process from here?

Brian M.: I suggest looking over the data and then learning more about the proposed analytical approaches, then identifying which approaches will be suitable. I’ll forward papers I have on N-Mixture models. [Also forwarding occupancy and removal papers].

Brian I.: UGA is looking to hire a quantitative bird person, but not sure if that will match this timeline.

Brian M.: If needed we can probably chat with Terri Donovan at University of Vermont. She was my post-doc advisor and is still active in this sort of modeling.